

# Entwicklung eines modular aufgebauten Programmsystems zur optimalen Erfassung, Speicherung und Auswertung von Daten bei klinischen Arzneimittelprüfungen\*)

## 3. Mitteilung: Datenstrukturen und Datenmodell\*\*)

R. Gordon, E. Hartmann, G. Laudahn und H. Röpke

### Zusammenfassung

*Die Inputstruktur für statistische Standardalgorithmen orientiert sich zweckmäßigerweise an der Grundgesamtheit des Analysenmodells. Die Primärstruktur von Daten aus klinischen Prüfungen orientiert sich dagegen weder am Prüfplan noch an der Art der geplanten statistischen Analysen, sondern an der Datenerfassung in der Klinik. Mit der hier vorgestellten Data Definition Language (DDL) können die zugrundeliegenden Strukturen bereits bei der Vergabe von Namen für Daten berücksichtigt werden. Die DDL bildet so die Basis für prüfplan-spezifische Umstrukturierungsmöglichkeiten.*

### Summary

*The most convenient input structure for standard statistical algorithms is based upon the population of the analytical model. The primary structure of clinical trial data, however, is based neither upon the trial plan, nor upon the analyses to be carried out, but upon the process of data acquisition in the clinic. With the Data Definition Language (DDL) presented here, underlying structures can be accounted for during the process of defining names for the data. The DDL thus serves as a basis for restructuring possibilities which are compatible with the trial plan.*

### 1. Einleitung

In den ersten beiden Mitteilungen [1, 2] wurden die wissenschaftlichen Voraussetzungen und die Zielsetzung des Programmsystems beschrieben sowie das Konzept unter biometrischen und DV-technischen Aspekten vorgestellt. In der vorliegenden Mitteilung wird über die adäquate Berücksichtigung der Versuchspläne bei der Erstellung des Datenmodells und über die daraus resultierenden Möglichkeiten zur Umstrukturierung der Daten berichtet.

\*) Gefördert durch das Bundesministerium für Forschung und Technologie, Projekt Datenverarbeitung in der Medizin, Teilvorhaben DVM 352 (vorm. DVM 116 bzw. DVM 302).

\*\*) Herrn Hultsch, Universität Münster, danken wir für wertvolle Anregungen und Diskussionen zu diesem Thema.

### 2. Die Architektur des Systems

Das Programmsystem besteht aus einem Datenbanksystem und aus einer Methodenbank. Mit dem Datenbanksystem kann ein Anwender eine prüfungsspezifische Datenbank<sup>1)</sup> generieren und darin die Speicherung, das Retrieval und das Handling der Daten der Prüfung veranlassen. In der Methodenbank sind die für statistische Analysen benötigten Programme zusammengestellt, auf die der Anwender mittels einer speziellen Kommandosprache zugreifen kann.

Ist die Datenbank für eine Prüfung generiert, so erfolgen die Auswertungen nach dem in Abb. 1 dargestellten Schema. Jede vom Anwender spezifizierte Fragestellung besteht aus einem Datenbankzugriff und einem Methodenbankzugriff.

Der Datenbankzugriff veranlaßt die Bereitstellung der benötigten Daten in einem Arbeitsbereich (dem sog. Worksheet), während der Methodenbankzugriff die Bereitstellung der Auswertungsprogramme veranlaßt.

Das so gewonnene Worksheet muß zweierlei Anforderungen erfüllen: erstens muß seine Struktur mit der Inputstruktur der Programme übereinstimmen; zweitens muß sein Inhalt die Durchführung sinnvoller Analysen ermöglichen. Beides muß in einem flexiblen System vom Anwender überwacht werden. Die prüfplanbedingten Strukturen, die den Daten zugrundeliegen, werden bei der Datenbankgenerierung in einem relationalen Datenmodell<sup>2)</sup> erfaßt.

Die Primärstruktur der in klinischen Prüfungen gewonnenen Daten orientiert sich in erster Linie an der Zweckmäßigkeit der Datenerfassung in der Prüfklinik. Der üblichen Praxis in der medizinischen Dokumentation folgend, wurde daher der Proband als Einheit der Datenerfassung gewählt. Der Planung entsprechend sind außerdem für jeden Probanden dieselben Messungen durchzuführen, so daß die Datenmodelle klinischer Prüfungen dem Typ nach Relationen sind, mit einer Zeile je Proband und einer Spalte je durchzuführende Messung. Das Aufstellen des Datenmodells erfolgt durch das Fest-

<sup>1)</sup> Die hier betrachteten Datenbanken sind – verglichen mit Krankenhausinformationssystemen oder den meisten kommerziellen Datenbanken – klein und einfach. Die Beschreibung des Systems wird aber durch die Verwendung der Sprechweise der Datenbanksysteme erheblich erleichtert. Näheres zu dieser Sprechweise siehe [3], [4].

<sup>2)</sup> Zu relationale Datenmodelle incl. der Begriffe »Relation« und »Attribut« vgl. [4] Seite 41 ff.



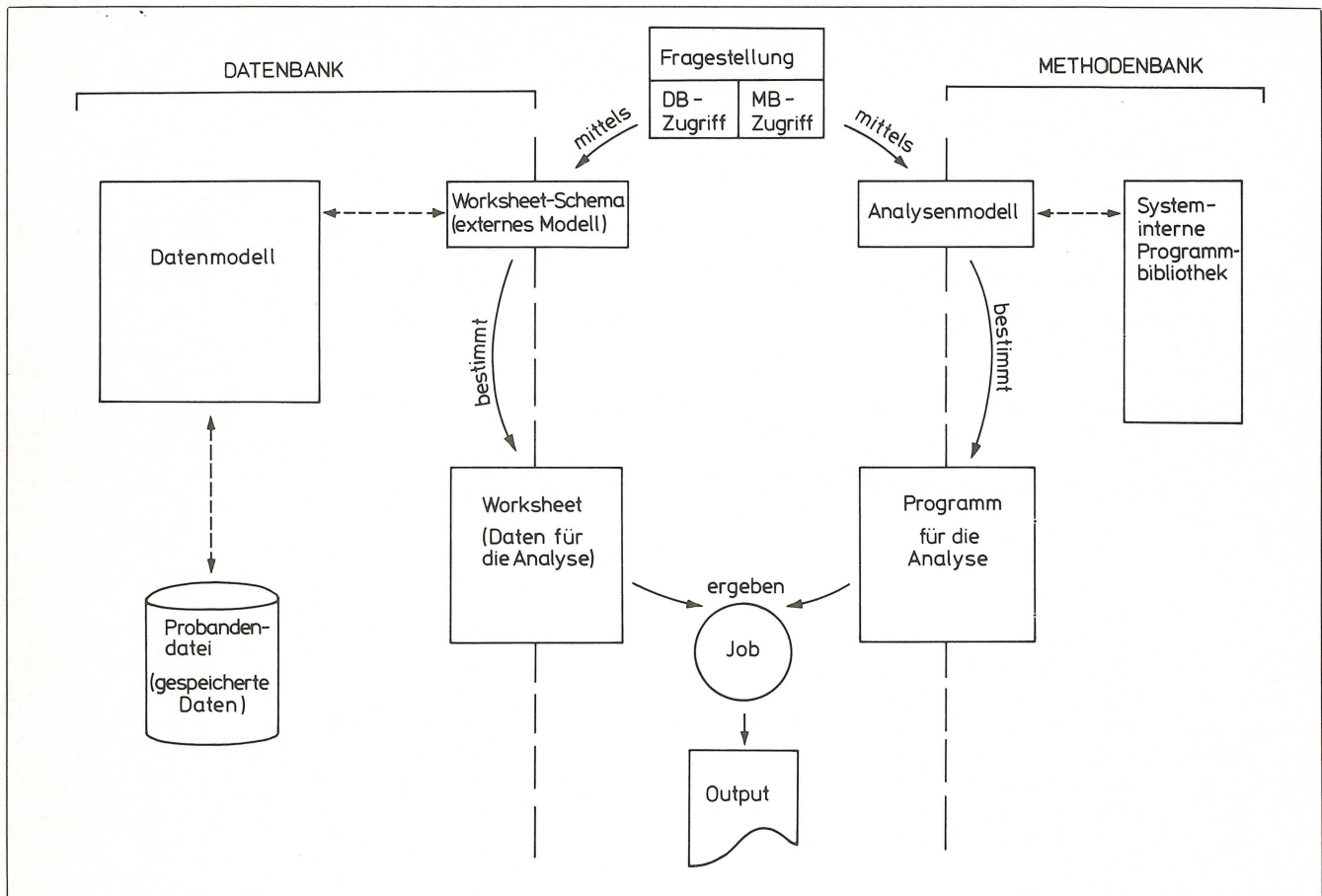


Abb. 1. Der allgemeine Ablauf einer Auswertung

legen von prüfungsspezifischen Attributen für die Spalten. Die »Data Definition Language« (DDL) ist so konstruiert, daß die prüfplanbedingten Strukturen in einer dem System zugänglichen Weise in den Attributen festgehalten werden können.

### 3. Erfassung der Datenstrukturen

Die mit Hilfe der DDL in den Attributen zu erfassenden Datenstrukturen resultieren aus dem Versuchsplan und aus dem Raster der Meßzeitpunkte (vgl. [2]). Die wesentliche Aufgabe des Versuchsplans liegt in der Definition der Versuchseinheiten sowie deren Zuordnung zu den zu prüfenden Behandlungen. Je nach Prüfplantyp sind dabei die Versuchseinheiten durch ihre Zugehörigkeit zu einer Teilmenge der folgenden Kategorien identifiziert:

- Prüfzentrum (Klinik)
- Block
- Proband
- Periode bzw. Seite.

Zur vollständigen Beschreibung der Struktur werden außerdem die Kategorien des Meßrasters innerhalb der Versuchseinheit:

- Meßzeitpunkt und
- Wiederholungsmessung je Meßzeitpunkt

benötigt [2].

Die Versuchseinheit braucht weder mit der Einheit der Datenerfassung noch mit der Analyseneinheit übereinzustim-

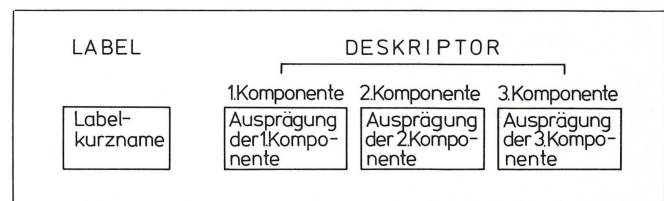
men, aus welcher die Grundgesamtheit der statistischen Analyse besteht.

Für die Zugehörigkeit der Analyseneinheit zu den Prüfzentren, Blöcken und Probanden liegen in dem probandenorientierten Datenmodell Attribute vor. Für die Strukturmerkmale Perioden, Seiten, Meßzeitpunkte und Wiederholungsmessungen können hingegen keine Attribute vorliegen. Dies sind die o.g. Informationen, die in den Attributen festgehalten werden müssen. Zu diesem Zweck werden die Attribute des Datenmodells aus folgenden zwei Bestandteilen aufgebaut (vgl. Abb. 2):

- Der Label beschreibt die Art der Messung.
- Der Deskriptor fixiert die Zuordnung des Attributs zu folgenden Strukturmerkmalen:
  - Periode oder Seite
  - Meßzeitpunkt und
  - Wiederholungsmessung.

Der Deskriptor besteht aus maximal drei Komponenten, von denen je eine zur Erfassung der Zuordnung zu einem dieser drei Strukturmerkmale vorgesehen ist. Ist ein Attribut einem dieser Strukturmerkmale nicht zuordenbar, so entfällt

Abb. 2. Form eines allgemeinen Attributs





die entsprechende Komponente im Deskriptor. Im Grenzfall, in dem das Attribut allein dem Probanden zuordenbar ist, entfällt der Deskriptor ganz.

Die Definition des Datenmodells erfolgt in drei Schritten:

- Im ersten Schritt wird angegeben, welche Komponenten für die gegebene Prüfung relevant sind.

- Im zweiten Schritt werden alle Ausprägungen jeder angegebenen Komponente festgelegt.

- Im dritten Schritt werden alle in der Prüfung auftretenden Labelkurznamen vereinbart. Gleichzeitig wird die Menge aller dazugehörenden Deskriptoren festgelegt. Diese Menge muß »balanciert«, d.h. ein kartesisches Produkt von Teilmengen der Komponenten<sup>3)</sup> sein. Sie wird durch die Angabe festgelegt, welche Komponentenausprägungen in jeder beteiligten Komponente auftreten dürfen.

Die beschriebene DDL spiegelt nicht nur die prüfplanbedingte Struktur klinischer Prüfungen wider, sondern gestattet zugleich eine mnemotechnisch günstige Wahl der Attribute. Beispielsweise ist direkt aus dem Attribut ersichtlich, ob zwei Messungen in derselben Periode durchgeführt wurden oder nicht.

#### 4. Festlegung des Worksheet-Schemas und Erstellung des Worksheets

Da die Programme der Methodenbank möglichst universell einsetzbar sein sollen, ist deren Inputstruktur an der statistischen Analyse, nicht aber am Prüfplan orientiert. Hier wurde als Struktur eine Relation mit einer Zeile je Element der analysenbedingten Grundgesamtheit und einer Spalte für jede in die statistische Analyse eingehende Variable gewählt.

Da die Programme der Methodenbank unabhängig von den Attributen des Datenmodells sind, wurde den Spalten dieser Relation eine feste Anordnung gegeben. Die bestimmungsgemäße Verwendung der Worksheetspalten im Auswertungsalgorithmus ergibt sich dann aus deren Reihenfolge.

Wenn die zu einer Analyse gehörende Grundgesamtheit aus Probanden besteht, so entspricht jede Zeile des Worksheets

<sup>3)</sup> Häufig wird sprachlich nicht zwischen einer Menge und dem Namen dieser Menge unterschieden. Entsprechend bedeutet Komponente entweder die Menge der Komponentenausprägungen oder den Namen dieser Menge (»Periode«, ...), je nach Kontext.

einer Zeile der gemäß dem Datenmodell aufgebauten Probandendatei. Die einfachste Form eines Worksheet-Schemas ist hiernach eine Liste von Attributen aus dem Datenmodell und das zugehörige Worksheet eine Projektion der Probandendatei.

Je nach Prüfplan können aber Analysen erforderlich werden, in denen die Elemente der Grundgesamtheit aus mehreren Probanden bestehen (z.B. matched pair) oder in denen ein Proband mehreren Elementen der Grundgesamtheit entspricht (z.B. change-over). Im ersten Fall wird jede Zeile des Worksheets aus mehreren Zeilen der Probandendatei gebildet, im zweiten entstehen aus jeder Zeile der Probandendatei mehrere Zeilen des Worksheets. In beiden Fällen muß der Systemanwender die durchzuführende Umstrukturierung zum Worksheet festlegen.

Die erforderlichen Umstrukturierungen lassen sich in drei Typen einteilen, deren Prinzip im Folgenden erläutert und veranschaulicht wird:

Typ 1: Der Prozeß soll zunächst am Beispiel einer Prüfung vom »change-over«-Typ demonstriert werden. Bei einem solchen Prüfplan sind Analysen durchzuführen, deren Grundgesamtheit aus »Probandenperioden« besteht. In das Datenmodell wurde hierzu die Komponente »Periode« eingeführt. Für die Erstellung eines auf die Analyse ausgerichteten Worksheet-Schemas müssen die benötigten Attribute modifiziert werden. Die Periode tritt nun nicht mehr als Komponente, sondern als Attribut auf, welches gemeinsam mit »Probandennummer« das Schlüsselattribut bildet. Aus Attributen mit der Komponente »Periode« werden durch Streichen der zugehörigen Ausprägung neue Attribute gebildet, die diese Komponente nicht mehr enthalten. Das zugehörige Worksheet enthält je Probandenperiode eine Zeile und wird den neuen Zuordnungen entsprechend mit den betreffenden Meßdaten aufgefüllt (vgl. Tab. 1).

Dieser Prozeß des »Abbaus« einer Komponente kann allgemein dadurch charakterisiert werden, daß diese Komponente aus den betreffenden Attributen eliminiert und zugleich als ein worksheetzeilen-definierendes Attribut eingeführt wird.

Typ 2: Als Beispiel hierfür soll eine Prüfung vom »matched-pair«-Typ angeführt werden. Hierbei werden Paare gebildet,

Tab. 1. Veranschaulichung des Abbaus einer Komponente

Probandennr.	Puls. Per 1	Puls. Per 2	Behandl. Per 1	Behandl. Per 2	Reihenf.	Alter	← Attribute im Datenmodell
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
21	70	63	Beh 1	Beh 2	1	28	
22	80	91	Beh 2	Beh 1	2	51	← Zeilen in der Probandendatei
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	

Probandennr.	Periode	Puls	Behandl.	Reihenf.	Alter	← Attribute nach Abbau von »Periode«
.	.	.	.	.	.	
.	.	.	.	.	.	
21	Per 1	70	Beh 1	1	28	
21	Per 2	63	Beh 2	1	28	
22	Per 1	80	Beh 2	2	51	← Worksheet-Zeilen, die durch Abbau entstehen
22	Per 2	91	Beh 1	2	51	
.	.	.	.	.	.	
.	.	.	.	.	.	

(Schlüsselattribut dick unterstrichen)



Tab. 2. Veranschaulichung des Aufbaus einer Komponente

Probandennr.	Paar	Behandlung	Größe	Gewicht	← Attribute im Datenmodell		
.	.	.	.	.	← Zeilen in der Probandendatei ←		
.	.	.	.	.			
17	7	Beh A	159	60.1			
18	2	Beh B	175	57			
.	.	.	.	.			
.	.	.	.	.			
27	2	Beh A	179	60.1			
.	.	.	.	.			
.	.	.	.	.			
40	7	Beh B	162	67.3			
.	.	.	.	.			
.	.	.	.	.			

Probandennr. Beh A	Probandennr. Beh B	Paar	Größe. Beh A	Größe. Beh B	Gewicht. Beh A	Gewicht. Beh B	← Attribute nach Aufbau von »Behandlung«
.	.	.	.	.	.	.	← Worksheetzeilen, die durch den Aufbau entstehen ←
.	.	.	.	.	.	.	
27	18	2	179	175	60.1	57	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
17	40	7	159	162	60.1	67.3	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	
.	.	.	.	.	.	.	

(Schlüsselattribut dick unterstrichen)

in denen die Probanden unterschiedliche Behandlungen erhalten. Bei einem solchen Prüfplan sind Analysen vorgesehen, deren Grundgesamtheit aus Paaren besteht. Solche Analysen beruhen beispielsweise auf Meßwertdifferenzen oder auf der Klassifikation von Ausprägungskombinationen je Paar. Im Datenmodell müssen dann Attribute »Paar« und »Behandlung« vorhanden sein, die anzeigen, zu welchem Paar ein Proband gehört und welche Behandlung er erhalten hat. Wesentlich ist hierbei, daß das zusammengesetzte Attribut »Paar«, »Behandlung« die Probandennummer als Schlüsselattribut ersetzen kann und daß in jedem Paar dieselbe Menge von Ausprägungen von »Behandlung« auftritt. In einem auf diese Analyse ausgerichteten Worksheet-Schema ist das elementare Attribut »Paar« das Schlüsselattribut, während »Behandlung« als Komponente anstatt als Attribut auftritt. Die Deskriptoren der benötigten Attribute werden um die entsprechende Ausprägung der neuen Komponente »Behandlung« ergänzt. Das zugehörige Worksheet enthält je Paar eine Zeile, die durch Zusammenfassen der entsprechenden Zeilen der Probandendatei entsteht (vgl. Tab. 2).

Dieser Prozeß des »Aufbaus« einer neuen Komponente kann allgemein dadurch charakterisiert werden, daß das betreffende Attribut aus dem Datenmodell eliminiert und dafür als Komponente aufgebaut wird. Die Attribute des Worksheet-Schemas entstehen aus denen des Datenmodells durch Ergänzen der Deskriptoren um die zutreffenden Ausprägungen

gen der neuen Komponente<sup>4</sup>). Der Aufbau einer neuen Komponente aus einem Attribut kann als eine Umkehrung des unter Typ 1 beschriebenen Prozesses angesehen werden. Er ist nur dann möglich, wenn das Attribut Teil eines zusammengesetzten identifizierenden Attributs ist, dessen vollständige Zusammensetzung dem System mitgeteilt wurde. Die Mitteilung kann in Form einer Randomisierungsliste geschehen.

Typ 3: Eine dritte Art der Umstrukturierung der Meßdaten ist speziell bei bestimmten Auswertungen von Change-over- oder Mehrseitenprüfungen von Nutzen, bei denen der Proband einem Versuchsblock entspricht (vgl. [2]).

In dem entsprechenden Datenmodell enthalten die relevanten Attribute, einschließlich der zur Kennzeichnung der Behandlung erforderlichen Attribute, die Komponente »Periode« oder »Seite«.

Aus der Zuordnung der Behandlungen zu den Versuchseinheiten und deren Zuordnung zu den Kategorien des Versuchsplans resultiert eine Redundanz in den zur Identifizierung der Versuchseinheiten verfügbaren Informationen. Diese Redundanz erlaubt z.B. im Change-over- bzw. Mehrseiten-

<sup>4</sup>) Diese Steuerungsmöglichkeiten wurden im Rahmen des Teilvorhabens DVM 352 konzipiert, jedoch ist die Realisierung aus Zeitgründen nicht im Rahmen des Teilvorhabens, wohl aber bei der Weiterentwicklung geplant.

Tab. 3. Veranschaulichung des Umbaus einer Komponente aus dem Datenmodell von Tab. 1

Probandennr.	Puls. Beh 1	Puls. Beh 2	Per. Beh 1	Per. Beh 2	← Attribute nach dem Umbau
.	.	.	.	.	← Worksheetzeilen, die durch den Umbau entstehen ←
.	.	.	.	.	
21	70	63	Per 1	Per 2	
22	91	80	Per 2	Per 1	
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	



versuch den Ersatz der Komponente »Periode« bzw. »Seite« durch »Behandlung«. Hierbei bleibt die Zeilendefinition, d.h. das Schlüsselattribut, des Datenmodells unverändert. Zur Erstellung des Worksheets müssen lediglich die Spalteninhalte zeilenabhängig umsortiert werden.

Diesen Prozeß kann man sich als einen Abbau der Komponente »Periode« bzw. »Seite« mit anschließendem Aufbau der Komponente »Behandlung« vorstellen.<sup>5)</sup>

Tabelle 3 veranschaulicht den Prozeß des Umbaus, wobei von dem in Tabelle 1 verwendeten Datenmodell ausgegangen wurde.

Die hier vorgestellte DDL dient als Bindeglied zwischen dem Prüfplan und den im System vorgesehenen Möglichkeiten der Worksheeterstellung. Das System bietet einerseits die erforderliche Flexibilität und Übersichtlichkeit des Data-handlings, erschwert aber andererseits die biometrisch inadäquate Vergrößerung von Stichprobenumfängen durch Umstrukturierung. Damit unterstützt es den Anwender bei der statistischen Interpretation der Ergebnisse klinischer Prüfungen.

<sup>5)</sup> Vgl. Fußnote S. 36

## Literatur

- [1] HARTMANN, E., LAUDAHN, G., LEHNERT, J., RÖPKE, H., ROSENFELD, E. (1974): Entwicklung eines modular aufgebauten Programmsystems zur optimalen Erfassung, Speicherung und Auswertung von Daten bei klinischen Arzneimittelprüfungen; 1. Mitteilung: Grundlagen und Ziele. EDV in Medizin und Biologie 5, 127–132.
- [2] DANNEHL, K., HARTMANN, E., LAUDAHN, G., LEHNERT, J., RÖPKE, H., ROSENFELD, E. (1976): Entwicklung eines modular aufgebauten Programmsystems zur optimalen Erfassung, Speicherung und Auswertung von Daten bei klinischen Arzneimittelprüfungen; 2. Mitteilung: System-Konzept. EDV in Medizin und Biologie 7, 96–103.
- [3] DATE, C. J. (1977): An Introduction to Database Systems, 2nd Ed. Addison-Wesley, Reading, Mass.
- [4] WEDEKIND, H. (1974): Datenbanksysteme I, Bibliographisches Institut, Zürich.

Anschrift für die Verfasser: Dr. Horst Röpke, Schering AG, Physikochemie und Informatik, Postfach 65 03 11, 1000 Berlin 65

EDV in Medizin und Biologie 10 (2), S. 37–43, ISSN 0300-8282  
© Eugen Ulmer GmbH & Co., Stuttgart; Gustav Fischer Verlag KG, Stuttgart

# Entwicklung eines modular aufgebauten Programmsystems\*) zur optimalen Erfassung, Speicherung und Auswertung von Daten bei klinischen Arzneimittelprüfungen

## 4. Mitteilung: Anwendungsbeispiel

R. Meister, H. Ellmer, U. Grünberg, E. Hartmann,  
J. Kaufmann, G. Laudahn, J. Lehnert, H. Röpke, E. Rosenfeld

### Zusammenfassung

ASPECT vereinigt alle Funktionen, die zur Auswertung einer klinischen Prüfung erforderlich sind. Der Ablauf einer Auswertung und Teile der Kommandosprache werden an einem Beispiel demonstriert. Für die systematische und mnemotechnische Datenbeschreibung können Begriffe des Versuchsplans und die Variablenamen verwendet werden. Die statistischen Verfahren von ASPECT berücksichtigen die speziellen Belange der klinischen Prüfung. Datenbeschreibung, Datahandling und Auswertung wurden aufeinander abgestimmt, so daß die Befehlsfolgen übersichtlich sind.

### Summary

ASPECT unites all functions necessary for the evaluation of a clinical trial. The course of the evaluation and parts of the com-

mand language are demonstrated on an example. The terms of the experimental design and the names of the variables can be used for systematic data description using mnemonics. The statistical procedures of ASPECT were selected to fulfil the needs of clinical trials. The commands for the data description, data handling and evaluation procedures are harmonized, so that they are easily read.

### 1. Einleitung

In den ersten drei Mitteilungen (6, 2, 5) wurden Zielsetzung und Konzept des Projektes vorgestellt. Mit dieser 4. Mitteilung werden die Ergebnisse an einem Beispiel erläutert. Dabei richtet sich das Hauptaugenmerk auf die Anwendung, nicht dagegen auf die DV-technischen Gesichtspunkte.

Seit Beginn der Arbeiten an dem System ASPECT bestand das Mitarbeiterteam, das gleichzeitig mit der Planung und Auswertung klinischer Arzneimittelprüfungen beschäftigt war, aus Medizinern, Statistikern und Informatikern. So konnten die erarbeiteten Anforderungen an die Datenbeschreibung, das Datahandling und die statistischen Analysen ständig in der Praxis überprüft werden.

Das in (2) vorgestellte Konzept wurde inzwischen weitgehend realisiert. Ergebnis der Arbeiten zur Datenbeschreibung

\*) Das Programmsystem wurde inzwischen »ASPECT« benannt (ASPECT – Approach to Systematic Planning and Evaluation of Clinical Trials). Es wurde durch das Bundesministerium für Forschung und Technologie gefördert, und zwar im Rahmen des Projektes Datenverarbeitung in der Medizin, Teilvorhaben DVM 352 (vorm. DVM 116 bzw. 302).



ist eine einfache Kommandosprache zur Erzeugung prüfungs-spezifischer Datenbanken. Die Bezeichnungen für Variable können dabei mnemotechnisch gewählt werden. Der Versuchsplan – z. B. Mehrperioden Change-Over, zeitlicher Verlauf, Mehrfachmessung – wird Bestandteil der zusammengesetzten Namen (5). Die weitere Bearbeitung wird wesentlich vereinfacht, da diese Definitionen für das Datahandling und für alle statistischen Analysen verwendet werden können. Durch Simulation von Problemanalysen konnte die Vielfalt klinischer Prüfungen auf sechs häufig auftretende Prüfplantypen (2) reduziert werden. Die daraus ableitbaren Datenstrukturen treten in der Praxis nicht bei allen Variablen der jeweiligen Prüfplantypen auf. In der Datenbeschreibung ist deshalb die Kopplung zum Prüfplantyp flexibel gestaltet worden. Die ausgewählten Methoden ermöglichen den größten Teil der erforderlichen statistischen Auswertungen. Über eine externe Datenschnittstelle können auch Spezialverfahren im Anschluß angewendet werden.

Es sollte die Durchführung kontrollierter klinischer Therapiestudien, wie sie in (3) gefordert wird, durch eine Planungs- und Auswertungshilfe unterstützt werden. Deshalb kann und wird ASPECT weder mit großen Datenbanksystemen noch mit umfangreichen Statistikprogrammpaketen konkurrieren. Das Hauptanliegen besteht in Datenschutz, in Datensicherheit sowie in Übersichtlichkeit und Überprüfbarkeit der Analysen durch die Integration aller benötigten Teilfunktionen in ein System.

## 2. Überblick

Bevor die Bestandteile von ASPECT im einzelnen beschrieben werden, soll die Einordnung des Systems in den Ablauf der Pla-

nung und Auswertung einer klinischen Prüfung erläutert werden.

Abbildung 1 stellt den Ablauf einer klinischen Prüfung von der allgemeinen Zielsetzung bis zum Ergebnisbericht dar. Der Einsatz von ASPECT beginnt nach der Datengewinnung und endet mit den Ergebnissen der erforderlichen Auswertungen. Dabei sind alle zur Auswertung notwendigen Aktivitäten innerhalb des Systems möglich und werden für jede Prüfung in einem gesonderten Logbuch dokumentiert. Die Berücksichtigung des Versuchsplans und die Systematik der Datenbeschreibung von ASPECT kann jedoch schon in der Planungsphase hilfreich für die Entwicklung der Prüfstrategie und der Prüfbogen sein.

In dieser Arbeit sollen der Aufbau und die Möglichkeiten des Systems anhand eines praxisnahen Beispiels erläutert werden. Als Versuchsplan wurde ein 2-Perioden-Change-Over-Design gewählt. Es wurde ein verkürzter Prüfbogen konstruiert, der einige Variable enthält, welche für die Indikation Hypertonie charakteristisch sind. Die Datenbeschreibung, das Datahandling und die statistischen Analysen von ASPECT werden exemplarisch im Rahmen einer möglichen Auswertung dieser Prüfung vorgestellt.

## 3. Die Datenbank

Bei jeder klinischen Prüfung werden Messungen, Beobachtungen und Erhebungen von Merkmalen durchgeführt. In diesem Abschnitt soll erklärt werden, wie eine Datenbank, welche die Realisierungen aller erhobenen Merkmale je Patient enthält, mit Hilfe der Data Definition Language (DDL) aufgebaut werden kann.

### 3.1. Datenbeschreibung

Abbildung 2 stellt den Prüfbogen dar, der in der klinischen Prüfung SH ABC 14 eines neuen Antihypertonicums gegen eine Standardtherapie als Erfassungsbeleg verwendet wurde. In der Anamnese finden sich die Variablen Alter, Geschlecht, Gewicht und Glucosekonzentration. Für Alter, Geschlecht und Gewicht gibt es nur eine Angabe je Patient. Dagegen wird bei der Glucosekonzentration dasselbe Merkmal zu mehreren Zeitpunkten bestimmt. Um die Variablen eindeutig zu definieren, müssen also Merkmalsname und Meßpunkt genannt werden.

Der Glucosetoleranztest würde in dieser Studie als Ausschlusskriterium bei Patienten mit Diabetesverdacht dienen. Die übrigen Variablen auf dem Prüfbogen sind in der Sprechweise der Biometrie die Zielgrößen der Studie.

Die blutdrucksenkende Wirkung der neuen Substanz im Vergleich zum Stan-

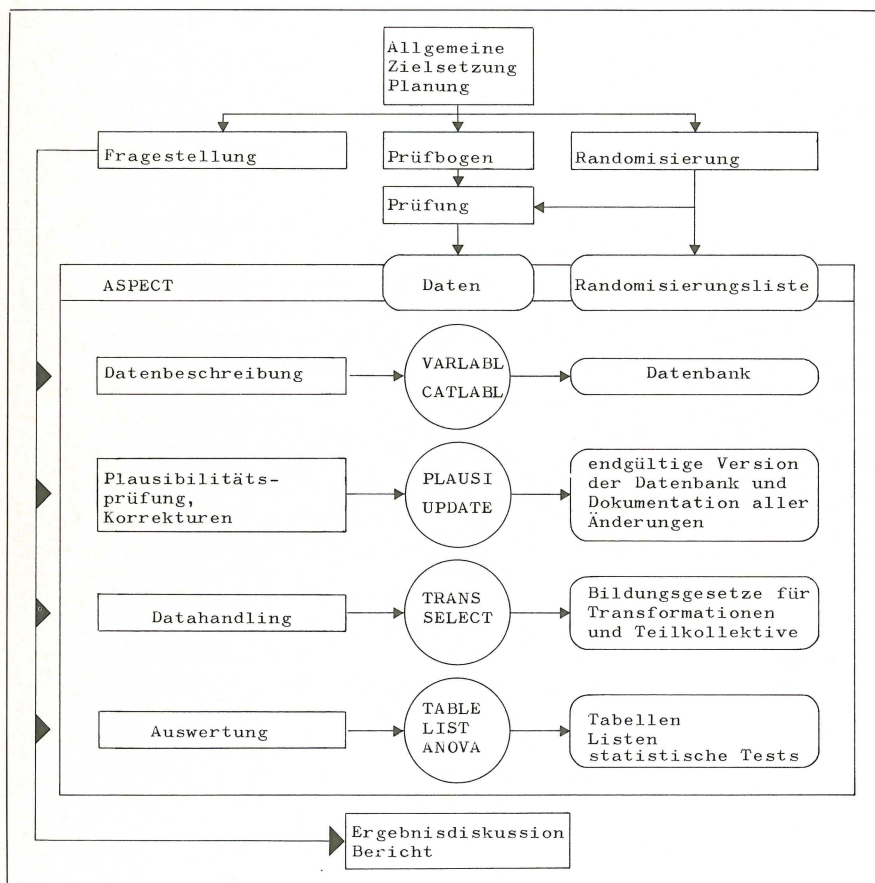


Abb. 1 (links). Die Stellung von ASPECT im Ablauf der Planung und Auswertung

Abb. 2 (Seite 39). Verkürzter Prüfbogen SH ABC 14



<b>SH ABC 14</b>			Patienten-Nr.	Prüfbeginn	Alter (Jahre)	Geschlecht	Gewicht (kg)	
<div style="display: flex; justify-content: space-between;"> <div> <input type="text" value="0"/> <input type="text" value="1"/> KA 1 2 </div> <div> <input type="text" value="1"/> <input type="text" value="4"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 3 4 5 6 7 8 </div> </div>			<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 9 10 11	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> Tag Monat Jahr 12 13 14 15 16 17	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 18 19	1 = männlich 2 = weiblich <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 20	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 21 22 23	
			0 Minuten	60 Minuten	120 Minuten			
Glukosekonzentration m mol/l			<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 24 25 26	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 27 28 29	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 30 31 32			
			<input type="text" value="0"/> <input type="text" value="2"/> KA 3-11 wie KA 01 12-15 = leer		<input type="text" value="0"/> <input type="text" value="3"/> KA 3-11 wie KA 01 12-15 = leer			
<b>Befunde</b>			Vorphase	1. Periode nach		2. Periode nach		
				1 Woche	2 Wochen	1 Woche	2 Wochen	
Blutdruck (mmHg)	1. Messung liegen	syst.	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 33 34 35	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 16 17 18	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 33 34 35	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 16 17 18	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 33 34 35	
			diast.	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 36 37 38	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 19 20 21	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 36 37 38	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 19 20 21	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 36 37 38
		2. Messung liegen		syst.	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 39 40 41	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 22 23 24	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 39 40 41	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 22 23 24
			diast.		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 42 43 44	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 25 26 27	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 42 43 44	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 25 26 27
	<b>Nebenerscheinungen</b>							
	Kopfschmerz		1 = nein 2 = ja	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 45	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 28	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 45	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 28	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 45
	Schwindel			<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 46	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 29	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 46	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 29	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 46
	Müdigkeit			<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 47	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 30	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 47	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 30	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 47
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> sonstige 50 51		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 48		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 31	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 48	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 31	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 48	
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> sonstige 52 53		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 49		<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 32	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 49	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 32	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 49	
54-80 = leer								
50-80 = leer								
Die Prüfung wird		1 = beendet 2 = abgebrochen <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 50		am, Datum	<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> Tag Monat Jahr 51 52 53 54 55 56	wenn abgebrochen, wegen	1 = Nebenerscheinungen 2 = andere Gründe <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 57	
wenn Nebenerscheinungen oder andere Gründe, bitte angeben							<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 58 59 60 61 <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> 62 63 64 65 66-80 = leer	
Datum <span style="float: right;">Unterschrift</span>								



dard sowie Art und Häufigkeit möglicher Nebenwirkungen sollten ermittelt werden.

Im Beispiel wurde als Versuchsplan ein 2-Perioden-Change-Over-Design gewählt. Dabei erhielt die eine Hälfte der Patienten das neue Antihypertonikum in der ersten und das Standardpräparat in der zweiten Therapieperiode, während der anderen Hälfte der Patienten die Präparate in umgekehrter Reihenfolge verabreicht wurden. Wie bei jeder klinischen Arzneimittelprüfung muß auch hier die zufällige Zuteilung der Patienten zu den Gruppen vorausgesetzt werden.

Durch die Wahl eines bestimmten Designs ergeben sich prüfplanttypische Datenstrukturen (2).

Auf dem Prüfbogen liegen identische Meßpunkttraster in den beiden aufeinanderfolgenden Therapieperioden vor. Eine Ausnahme bildet die Vorphase, welche in diesem Falle nur der Kollektivbeschreibung dient. Die möglichen Nebenwirkungen wurden einmal in der Vorphase und je zweimal in wöchentlichem Abstand in beiden Therapieperioden erhoben. Der systolische und diastolische Blutdruck im Liegen wurde jeweils doppelt bestimmt. Bei den Zielgrößen muß genau wie bei der Glucosekonzentration der Merkmalsname<sup>1)</sup> und Meßpunkt – hier Periode und Meßzeit –, beim Blutdruck zusätzlich die Meßwiederholung (Abbildung 2) angegeben werden. Für die Computerauswertungen müssen alle Angaben auf Datenträger übertragbar sein. Für ASPECT werden die Eintragungen des Prüfbogens auf Lochkarten übertragen. Auf dem Prüfbogen befindet sich zum Beispiel die Altersangabe auf Karte 1 in den Spalten 18–19. Bei nominalen Größen, wie hier dem Geschlecht, sind den möglichen Werten der Variablen bestimmte Ausprägungen zugeordnet. Auf dem Bogen bedeutet eine »1« auf

<sup>1)</sup> Die Systematik der Namensvergabe wird in (5) im Rahmen der relationalen Datenmodelle beschrieben. Dort werden für »Variable« bzw. »Merkmalsname« die in diesem Rahmen gebäuchlicheren Begriffe »Attribut« bzw. »Label« verwendet. Der »Meßpunkt« wird durch den Deskriptor festgehalten.

Karte 1 in Spalte 20 »männlich« und eine »2« an dieser Stelle »weiblich«. Mit Hilfe der Patientennummer, die in allen Karten vorkommen muß, erhält man nun einen vollständig definierten, einem Patienten eindeutig zugeordneten Datensatz.

Zur Übernahme des Prüfbogens in das System sind die Befehle in Abbildung 3 erforderlich. Bei den in Abbildung 3 unterstrichenen Begriffen handelt es sich um reservierte Worte. Jedes Kommando wird durch ein Semikolon beendet.

Wie man sieht, wird für das Alter im Kommando VARLABL der Name AGE und die Stellung im Kartenschema durch »Karte 1 Spalten 18–19« bestimmt. Diese Angaben finden sich bei allen Variablen, die vom Prüfbogen in das System übernommen werden sollen. Ist wie bei der Glucosekonzentration noch Information über den Meßzeitpunkt erforderlich, können die Meßzeiten im Kommando TIMES vereinbart werden. Die Bezeichnung GLUCOSE.MIN steht dann für die Messungen nach 0, 60 und 120 Minuten.

Bei komplizierten Meßpunkttrastern stehen bis zu drei Komponenten zur Verfügung (5). Für die Nebenwirkungen besteht das Raster in diesem Beispiel aus Perioden- und Wochenangaben. Mit der Abkürzung PER.WEEK werden die Meßpunkte PER1 W1, PER1 W2, ..., PER2 W2 zusammengefaßt. Dabei wird die Komponente PERiode bei der Übernahme der Randomisierungsliste in das System definiert. Die Übernahme der Randomisierung geschieht vor der Variablendefinition, so daß die Berücksichtigung prüfplantypischer Strukturen in der Datenbeschreibung garantiert ist. Für reine Wiederholungsmessungen steht im System die Komponente REPLICATION zur Verfügung. Das zugehörige Meßpunkttraster wird durch die Angabe der Zahl der Wiederholungen definiert.

Die Zuordnung von Texten zu möglichen Ausprägungen der Variablen geschieht im Kommando CATLABL hier gemeinsam für alle im Kommando VARLBSET zusammengefaßten Nebenwirkungen. Die im Klartext erfaßten sonstigen Begleitsymptome werden nach Prüfungsende mit einem Nebenwirkungsthesaurus verschlüsselt. Durch geeignete Transformationen stehen sie später im System rein technisch wie die geplant abgefragten Variablen zur Verfügung.

Abb. 3. DDL–Datenbeschreibung in ASPECT

[illegible]

### 3.2. Plausibilitätsprüfungen und Update-Funktionen

Für jede Auswertung muß die Überprüfung der gewonnenen Befunde an erster Stelle stehen. Vor allem bei umfangreichem Datenmaterial sind Fehler sowohl bei der Erfassung als auch bei der Übertragung auf Datenträger unvermeidlich. Die Plausibilitätsprüfungen können sich dabei auf einfache Vergleiche der Ausprägungen von Variablen mit zulässigen Werten oder auf logische Verknüpfung mehrerer Vergleiche beziehen.

In ASPECT können die durch Plausibilitätsprüfungen gefundenen fehler- oder lückenhaften Datensätze durch Update-Funktionen korrigiert, ergänzt und gelöscht werden. Ebenso ist das Hinzufügen und Ändern von Variablendefinitionen und Textinformationen möglich. Jeder Update-Vorgang wird in eine »Archiv-Datei« mit einer Kennzeichnung übernommen, um Manipulationen an den Ursprungsdaten festzuhalten. Das Update wird wie alle übrigen Abläufe in dem »Logbuch« dokumentiert.

»Archiv« und »Logbuch« ermöglichen die schrittweise Wiederholung der Arbeiten zum Aufbau der Datenbank.

Anhand des Prüfbogens (Abb. 2) soll eine mögliche Plausibilitätsprüfung dargestellt werden. Für die Auswertung soll garantiert sein, daß bei den verschlüsselten sonstigen Nebenwirkungen mindestens einmal die Angabe »ja« während der Therapiephasen vorkommt. Das Kommando zur Plausibilitätsprüfung lautet dann:



**PLAUSI**

ELSEKEY1 EL (1,2,3,4,5) IFFI EXT PER EXT WEEK  
 ELSE1.PER.WEEK EQ 2 ;

In Worten: Genau dann, wenn der Schlüssel<sup>2)</sup> für sonstige Nebenwirkungen einen der Werte 1, 2, ..., 5 annimmt, existiert mindestens eine Periode und eine Woche, in der eine 2 ( $\triangleq$  ja) als Angabe vorkommt.

Nach Abschluß der Plausibilitätsprüfungen und des Update ist der Aufbau der Datenbank für eine Prüfung beendet.

**3.3. Datahandling**

Alle Aktivitäten, die zwischen dem Aufbau der Datenbank und der eigentlichen Auswertung liegen, werden hier als Datahandling bezeichnet. Typische Aufgaben des Datahandlings sind die Definition von Teilkollektiven und Transformationen im weiten Sinne. Zur Bildung von Teilkollektiven können beliebige Boole'sche Ausdrücke verwendet werden. Für die Transformationen läßt sich die Form der Variablendefinition ausnutzen. Insbesondere können bei Meßpunkttrastern Parametrisierungen vorgenommen werden. So lassen sich z. B. in einfacher Weise Variable wie Maximum über die Zeit, Mittelwert über Wiederholungen etc. gewinnen.

Am Beispiel wird gezeigt, wie die verschlüsselten sonstigen Nebenwirkungen »entschlüsselt« werden können, und eine Parametrisierung angegeben.

Angenommen, für »WEAKNESS« wäre der Schlüssel 5 unter sonstigen Nebenwirkungen vergeben, dann erhält man WEAKNESS mit demselben Meßpunkttraster wie die direkt abgefragten Nebenwirkungen durch folgende Transformation:

```
TRANS WEAKNESS.PER.WEEK = 1 ;
TRANS WEAKNESS.PER.WEEK
      = ELSE1.PER.WEEK IF ELSEKEY1 EQ5;
TRANS WEAKNESS.PER.WEEK
      = ELSE2.PER.WEEK IF ELSEKEY2 EQ5;
```

Die Bildung des Mittelwertes aus den Wiederholungsmessungen erreicht man durch

```
TRANS MRDIAS.T.PER.WEEK =
      MEAN (RRDIAS.T.PER.WEEK.REP
      P = REP);
```

Bei numerischen Variablen können mehrere Ausprägungen als fehlende Werte festgelegt werden. Systemintern werden diese Ausprägungen alle gleich behandelt. So konnte die automatische Behandlung fehlender Werte in Plausibilitätsprüfungen, Transformationen und Auswertungen einheitlich geregelt werden.

**4. Die statistischen Analysen in ASPECT**

Dieser Abschnitt beschreibt die in ASPECT verfügbaren statistischen Verfahren. Die Datenschnittstelle des Systems und methodische Fragen werden diskutiert.

**4.1. Daten**

Wie in (5) berichtet, orientiert sich die Primärstruktur von Daten aus klinischen Prüfungen weder vollständig am Prüfplan noch an der Art der geplanten Analysen, sondern im wesentli-

chen an der Datenerfassung in der Klinik. Mit Hilfe der in Abschnitt 3 beschriebenen DDL ist eine systematische, durch Prüfplan und Meßpunkttraster vorgegebene Variablendefinition möglich. Diese Systematik ist aber nicht immer für statistische Analysen geeignet.

Für den Statistiker sind die Meßdaten Realisierungen univariater oder multivariater Stichproben von Zufallsvariablen. Die Wahrscheinlichkeitsverteilung dieser Zufallsvariablen wird dabei im wesentlichen durch die im Experiment festgesetzten Einflußgrößen bestimmt. Notiert man alle zu analysierenden Meßdaten mitsamt allen zugehörigen Einflußgrößen je Stichprobenelement in einer Zeile und schreibt alle Zeilen untereinander, erhält man eine Datenmatrix, die alle Informationen über die Stichprobe enthält. Solche Datenmatrizen werden im Prinzip von allen Statistik-Systemen verwendet und sind in den Systemen MINITAB, OMNITAB (8) und RUMMAGE (10) als »worksheet« eingeführt. Das worksheet ermöglicht die Übergabe aller erforderlichen Informationen über die Daten und ist deshalb als universelle Datenschnittstelle systemintern und -extern einsetzbar.

In ASPECT erfolgt die Definition eines worksheets durch Angaben über die Stichprobenelemente und die Variablen, wodurch die Zeilen bzw. Spalten der Datenmatrix identifiziert werden (5). Damit ist die Worksheetstruktur nicht an die probandenorientierte Datenerfassung gebunden.

**4.2. Methoden**

Die Auswahl der statistischen Methoden orientiert sich an den Fragestellungen der klinischen Prüfungen. Für die qualitativen Befunde wurden analysenorientierte Tabellenprogramme entwickelt. Zur Analyse quantitativer Merkmale stehen parametrische und nichtparametrische Verfahren zur Verfügung.

Zur Dropout Analyse wurden Life Table Programme entwickelt. Darüber hinaus können Histogramme und graphische Darstellungen (Plots) erzeugt werden. Die Besonderheiten der Methoden werden einzeln erläutert.

**Häufigkeitsanalysen**

Die tabellarische Darstellung von Häufigkeitsverteilungen erfüllt bei der Auswertung eine Doppelfunktion. Die Abhängigkeit gewisser Merkmalsfrequenzen von Einflußgrößen soll untersucht und die Ergebnisse sollen in komprimierter Form übersichtlich dargestellt werden. Die Erfüllung dieser Doppelfunktion dient in (4) zur Beurteilung von Tabellenprogrammen.

Zur Lesbarkeit sind die Analysen und Tabellen in ASPECT auf maximal vier Dimensionen beschränkt. Die graphische Darstellung wird durch spezielle Befehle festgelegt. Alle definierten Texte stehen zur Beschriftung automatisch zur Verfügung.

Zur sinnvollen statistischen Analyse wurde die Erzeugung von Randverteilungen frei wählbar gestaltet. Die Prozentuierung wird durch die Angabe von abhängigen Variablen gesteuert.

Folgende Typen von Analysen wurden realisiert:

Kreuztabellen und graphisch zusammengesetzte Kreuztabellen für unverbundene Anordnungen sowie Symmetrietabellen für verbundene Anordnungen.

Anschließende Tests sind möglich.

<sup>2)</sup> Schlüssel  $\triangleq$  Zuordnung von Schlüsselzahlen zu Variablen.



## Lineare Modelle für parametrische Analysen

Zur Analyse quantitativer Merkmale sind lineare Modelle unersetzlich. Dies trifft in der experimentellen Forschung insbesondere für die Varianz- und Kovarianzanalyse zu. In der klinischen Forschung kann jedoch im Gegensatz zum Laborexperiment die Balanciertheit der zu analysierenden Daten nicht garantiert werden.<sup>3)</sup>

Die Bearbeitung dieser Probleme erfordert Erfahrung und spezielle Sachkenntnis bei der biometrischen Analyse und die Verfügbarkeit geeigneter Programme. Unserer Meinung nach gehört RUMMAGE (10) heute zu den wenigen Statistikprogrammen, mit denen alle zulässigen Analysen im unbalancierten Fall möglich sind (9). In Zusammenarbeit mit den Autoren G. R. Bryce und D. T. Scott wurden Teile von RUMMAGE modifiziert und in das System integriert.

Das Methodenspektrum umfaßt uni- und multivariate Varianz-, Kovarianz- und Regressionsanalysen, jeweils auch für unbalancierte Daten. Je Analyse wird nur ein Modell verwendet, da der eigentlichen Fragestellung klinischer Prüfungen nur hypothesengesteuerte und nicht datengesteuerte Analysen entsprechen. Die Möglichkeiten der Modellwahl sind in Tabelle 1 dargestellt (7). Hervorzuheben sind die Varianzanalysen für fixed, random und mixed models.

## Nichtparametrische Methoden

Der Flexibilität linearer Modelle mit vielfältigen Anwendungsmöglichkeiten steht bei den nichtparametrischen Methoden der Vorteil der Robustheit und schwächerer Voraussetzungen gegenüber. Die Erfahrungen mit dem Einsatz nichtparametrischer Methoden für große Stichprobenzahlen sind noch begrenzt; so wurden zunächst nur klassische Zwei- und Mehrstichprobenvergleiche für verbundene und unverbundene Anordnungen ausgewählt.

Für die Anwendung wird das Auftreten von Bindungen problematisch. Umfangreiche Untersuchungen der vorhandenen Literatur<sup>4)</sup> ergaben jedoch, daß schon bei mittleren Stichprobenumfängen die Ermittlung der exakten Verteilung der Test-

statistiken unter Bindungen einen unverhältnismäßig hohen Rechenaufwand erfordert. Auch sind keine einfachen Aussagen über die Zahl der zulässigen Bindungen und Verteilungskonvergenz der Teststatistiken möglich. Deshalb werden nur Angaben über die Zahl der Bindungen gemacht und die bindungskorrigierten Teststatistiken berechnet, deren asymptotische Verteilung bekannt ist. Es wurden der Kolmogorow-Smirnow-, der Wilcoxon signed rank- und der U-Test, der Kruskal Wallis- und der Friedman-Test realisiert.

Die verwendeten Algorithmen wurden von E. Hultsch (Universität Münster) entwickelt und in Zusammenarbeit in ASPECT integriert.

## Life Table Methoden

Zur Auswertung von Langzeitstudien ist eine korrekte Dropout Analyse unerlässlich. Aus der Bevölkerungsstatistik und der Epidemiologie stehen hierfür die Life Table Methoden zur Verfügung. Mit Hilfe dieser Methoden kann das Abbruchgeschehen in einer Prüfung analysiert werden. Ebenso ist die Untersuchung von Erfolgsraten in Abhängigkeit von der Zeit möglich. Die Vorteile des Einsatzes der Life Table Methoden gegenüber einer globalen Beurteilung zeigte Wolfrum (12) am Beispiel einer klinischen Prüfung von Kontrazeptiva.

Für ASPECT wurde Chiang's (1) Modell der konkurrierenden Risiken ausgewählt. Das Abbruchgeschehen kann damit übersichtlich und im Zeitverlauf dargestellt werden. Von Vorteil ist die mögliche Kumulation der Abbruchwahrscheinlichkeiten bis zum Endpunkt einer Studie und die Möglichkeit der hypothetischen Ausschaltung störender Abbruchgründe.

## 5. Auswertungsbeispiele

Die Einsatzmöglichkeiten der Auswertungsmethoden sollen für einige typische Fragestellungen vorgestellt werden.

1) Anlistung der Anamnesebefunde Alter, Gewicht, Geschlecht für jeden Probanden:

```
PRINT AGE SEX WEIGHT ;
```

Damit wird eine Liste mit den zugehörigen Überschriften erzeugt.

2) Erstellung einer Datei zur externen Weiterverarbeitung der Ergebnisse der Glucosetoleranztests:

```
FILE GLUCO.MIN ;
```

Das Kommando erzeugt eine Datei, die zur Weiterverarbeitung verwendet werden kann.

<sup>3)</sup> Zu diesem Problem bemerkt Searle (11): »... we are witnessing a great increase in the demand for analysis of unbalanced data, analysis which *cannot* be made merely by means of minor adjustments to traditional analysis of variance of balanced data. Indeed the situation is just the opposite: unbalanced data have their own analysis of variance techniques, and those for balanced data are merely special cases of the techniques for unbalanced data«.

<sup>4)</sup> Für die Durchführung der Untersuchungen danken wir Herrn W. Bensmann.

Tab. 1. Lineare Modelle (entnommen aus [7])

		Faktoriell		Cov.		Hierarch.		Regr.	
		Uni.	Mult.	Uni.	Mult.	Uni.	Mult.	Uni.	Mult.
Fixed	bal.	+	+	+	+	+	+		
	unbal.	+	+	+	+	+	+	+	+
Random	bal.	+	+	+	+	+	+		
	unbal.	+	+	+	+	+	+		
Mixed	bal.	+	+	+	+	+	+		
	unbal.	+	+	+	+	+	+		



3) Analyse des diastolischen Blutdrucks im Change-Over-Modell. In der üblichen Schreibweise der statistischen Literatur lautet ein Modell ohne Überhangeffekte:

$$Y_{ijk} = \mu + \alpha_i^R + \alpha_j^B + \alpha_{ij}^{R \times B} + p_{ik}^P + e_{ijk}$$

Hierbei bedeuten:  $\mu$  allgemeiner Mittelwert;  $\alpha_i^R$  Reihenfolgeeffekt,  $i = 1, 2$  fixed;  $\alpha_j^B$  Behandlungseffekt,  $j = 1, 2$  fixed;  $\alpha_{ij}^{R \times B}$  Wechselwirkung<sup>5)</sup>;  $p_{ik}^P$  Probandeneffekt,  $i = 1, 2$ ,  $k = 1, \dots, K$ , hierarchisch, random;  $Y_{ijk}$ ,  $e_{ijk}$  Meßwert und Fehler am  $k$ -ten Probanden in Reihenfolge  $i$  unter Behandlung  $j$ .

Der Analysenauftrag in ASPECT lautet:

```
ANOVA
DELCOM PER
DEPE MRDIAS.T.PER.WEEK MRSYST.PER.WEEK
FACTOR SEQINDEX TMTINDEX.PER PNINDEX
MODEL Y(I,J,K) = S(I)+T(J)+ST(I,J)+P(I,K)+E
FIXED S 2
FIXED T 2
RANDOM P
Optionen
END;
```

Durch diese Kommandos wird die übliche Varianztabelle mit den »expected meansquare coefficients« und zusätzlichen, durch Optionen steuerbaren Output erzeugt. Dieser Output entsteht für den gemittelten diastolischen und systolischen Blutdruck jeweils für die erste und zweite Woche.

4) Alters- und Geschlechtsverteilung

```
TABLE AGECL BY SEX
DEPE AGECL ;
```

Zur Auswertung wird die Transformation der Variablen AGE in Altersklassen AGECL benötigt. Je Geschlecht werden in der Tabelle die absoluten und relativen Häufigkeiten in den Altersklassen ausgedruckt.

5) Darstellung des Verlaufs der Nebenwirkungen über alle Perioden und Wochen

```
COMTABLE
SIDEFF.PER.WEEK / SEQ
COM VARLBSET SIDEFF BY PERIODS
TIMES ;
```

Je Reihenfolge (SEQ) wird eine Tabelle erzeugt, so daß die Behandlungen je Periode zugeordnet werden können. Durch die Angabe der Bezeichnung SIDEFF.PER.WEEK wird je Nebenwirkung, Periode und Woche das Tabellenbild für eine Einzeltabelle definiert.

Die Angaben nach COMBINED ergeben die Zusammensetzung der Gesamttabelle. In der Vorspalte werden die einzelnen Nebenwirkungen mit ihren Ausprägungen aufgelistet, in der Kopfzeile die Kombination aus Perioden und Wochen.

Die Erzeugung verschiedener Tabellentypen wird in einem Tabellenkatalog zum Benutzermanual beschrieben werden.

#### DV-technische Angaben

Das Programmsystem ASPECT läuft z. Z. auf einer Anlage CHB 66/40 mit 256 KW<sup>6)</sup> unter dem Betriebssystem GCOS<sup>7)</sup>.

<sup>5)</sup> Diese Wechselwirkung entspricht genau dem Periodeneffekt in einer alternativen Modellformulierung.

<sup>6)</sup> 1 KW = 1024 Worte; 1 Wort = 36 Bit

<sup>7)</sup> General Comprehensive Operating Supervisor

Der Einsatz erfolgt abhängig von der Aufgabenstellung im Batch-Processing oder Remote Job Entry. Die zur Systemunterstützung notwendigen Dateien können mit TEXT-EDITOR-Funktionen verwaltet werden.

Der Hauptspeicherbedarf beträgt ca. 80 KW im Overlay, sonst ca. 180 KW.

Die Größe der erforderlichen Dateibereiche ist im wesentlichen abhängig von der Anzahl der Probanden und Variablen. Bei einem Testbeispiel wurden für 41 Probanden mit je 233 Variablen einschließlich aller Text- und Steuerinformationen ca. 32 KW Plattenplatz benötigt. Sämtliche Programme sind in ANSI-FORTRAN geschrieben. Es wurden zwei spezielle CHB-System-Routinen benötigt: für die Null-Blank-Unterscheidung (Missing Values) und die FORTRAN-Formatüberschreitung; ansonsten wird nur Standardsystemsoftware benötigt.

Die Umstellung des Programmpaketes auf eine Siemens 7.755 mit BS 2000 ist in Vorbereitung.

#### Literaturverzeichnis

- (1) CHIANG, C. L. (1968): Introduction to Stochastic Processes in Biostatistics. John Wiley & Sons, New York
- (2) DANNEHL, K., HARTMANN, E., LAUDAHN, G., LEHNERT, J., RÖPKE, H., ROSENFELD, E. (1976): Entwicklung eines modular aufgebauten Programmsystems zur optimalen Erfassung, Speicherung und Auswertung von Daten bei klinischen Arzneimittelprüfungen. 2. Mitteilung: System-Konzept. EDV in Medizin und Biologie 7, 96–103
- (3) FINK, H., (Mitarb.), JESDINSKY, H. J. (Hrsg.) (1978): Memorandum zur Planung und Durchführung kontrollierter klinischer Therapiestudien, Hrsg. H. J. Jesdinsky, Schriftenreihe der GMDS 1
- (4) FRANCIS, I. (1976): Languages and Programs for Tabulating Data from Surveys; Handout at the Symposium on the Interface of Computer Science and Statistics
- (5) GORDON, R., HARTMANN, E., LAUDAHN, G., RÖPKE, H. (1979): Entwicklung eines modular aufgebauten Programmsystems zur optimalen Erfassung, Speicherung und Auswertung von Daten bei klinischen Arzneimittelprüfungen. 3. Mitteilung: Datenstrukturen und Datenmodell. EDV in Medizin und Biologie, Heft 2/79
- (6) HARTMANN, E., LAUDAHN, G., LEHNERT, J., RÖPKE, H., ROSENFELD, E. (1974): Entwicklung eines modular aufgebauten Programmsystems zur optimalen Erfassung, Speicherung und Auswertung von Daten bei klinischen Arzneimittelprüfungen. 1. Mitteilung: Grundlagen und Ziele. EDV in Medizin und Biologie 5, 127–132
- (7) HARTMANN, E., KAUFMANN, J. (1975): MAD – Ein Programm zur Auswertung linearer Modelle. Statistical Software Newsletters 2, 13
- (8) JOWETT, D., CHAMBERLAIN, R. L., MEXAS, A. G. (1972): OMNITAB – A Simple Language for Statistical Computations. J. Statist. Comput. Simul. 1, 129–147
- (9) KAUFMANN, J. (1979): Probleme der Anwendung mehrfaktorieller Varianzanalysen mit ungleichen Zellohäufigkeiten, Vortragsmanuskript. Kolloquium Biometrie in der experimentellen Forschung, Freie Universität Berlin
- (10) SCOTT, D. T., BRYCE, G. R., CARTER, M. W. (1975): RUMMAGE, A General Linear Models Data Analysis System. Manual, Brigham Young University, Provo/Utah.
- (11) SEARLE, S. R. (1971): Linear Models. John Wiley and Sons, New York, 139.
- (12) WOLFRUM, C. (1977): The Life Table according to Chiang as a Method for the Description and Assessment of the Efficacy of Oral Contraceptives. Meth. Inform. Med. 16, 176–181



# Untersuchungen zur Realisierung einer rechnergestützten Diagnosehilfe auf der Grundlage des langsamen akustisch evozierten Potentials

E. Stürzebecher, H. Wagner, H.-J. Gerhardt<sup>1</sup>, K.-D. Wernecke<sup>2</sup>, S. Unger<sup>3</sup>

## Zusammenfassung

Ausgangspunkt für die Untersuchungen war die Hypothese, daß der AEP-Zeitverlauf mehr Information über den Funktionszustand des akustischen Sinneskanals enthält, als die übliche Ausmessung von 3 Werten ( $N_1P_2$ -Amplitude,  $N_1$ - und  $P_2$ -Latenzen) liefert. Eine Prüfung dieser Frage erfolgte mit Hilfe multivariater Klassifizierungsverfahren (Diskriminanzanalyse, lernfähiger Trennebenenklassifikator) auf der Grundlage von 3 Modellstichproben: 1. Normalpersonen (76 Personen), 2. Patienten mit Lärmherschäden (35 Patienten), 3. Hörsturzpatienten (16 Patienten). Zur Beurteilung der Leistungsfähigkeit der mit diesen relativ kleinen Stichproben erstellten AEP-Klassifikatoren wurden die als R-Methode (zu optimistische Schätzung) und  $\pi$ -Methode (zu pessimistische Schätzung) bekannten Fehlerschätzverfahren eingesetzt. Die Diskrepanz zwischen den Ergebnissen beider Methoden ist relativ groß. Durch gewichtete Mittelung der Schätzwerte von R- und  $\pi$ -Methode wird eine weitere Eingrenzung des bei Neuklassifikationen zu erwartenden Fehlers erreicht. Die Fehlerrate eines auf der Grundlage der vorliegenden kleinen Stichproben erstellten Klassifikators liegt danach mit großer Wahrscheinlichkeit zwischen 22 % und 27 %. Bei einer für den praktischen Einsatz des Klassifikators erforderlichen Vergrößerung der Trainingsstichproben verschiebt sich der angebbare Fehlerbereich zu kleineren Werten. Vorteile bei der Anwendung des lernfähigen Trennebenenklassifikators gegenüber der Diskriminanzanalyse konnten für die vorliegenden Daten nicht nachgewiesen werden.

## Summary

Today there is no doubt that also specific parts of the cortex are involved in generating the slow AEP. Therefore it is likely that the AEP-time behaviour contains more for differential diagnostic purposes useful information than it is obtained by the usual measurement of only 3 values ( $N_1P_2$ -amplitude, 2 latencies). A test of this hypothesis was carried out by means of 2 different classification rules (discriminant analysis and a trainable classifier) on the basis of 3 samples: 1. normal hearing persons (76 persons), 2. patients with noise induced hearing loss (35 patients), 3. patients with sudden deafness (16 patients). To judge the performance (estimating the probability of misclassification)

of the classifiers being established on the given small samples we used 2 estimation rules, known as R-method (over optimistic) and  $\pi$ -method (over pessimistic). There is a relatively great discrepancy between the results of both methods. By a weighted averaging of the R- and  $\pi$ -method estimates a smaller range can be localized, where the actual rate of misclassification is to be expected. Using 20 variables the in this way estimated error of a classifier being established on the given small samples amounts to 22 % – 27 %. For practical employment greater samples are required to reduce the probability of misclassification. For the present data set there are no essential differences between the results of the discriminant analysis and the trainable classifier to be found.

## 1. Einleitung

Die langsamen akustisch evozierten Potentiale (AEP) dienen bis heute in der otologischen Diagnostik im wesentlichen nur zur objektiven Schwellenbestimmung. Hierfür sowie für die aus der Literatur bekannten Ansätze einer überschwelligigen Diagnostik auf AEP-Basis werden meistens nur 3 Werte ( $N_1P_2$ -Amplitude,  $N_1$ - und  $P_2$ -Latenz, siehe Abb. 1) an dem durch Mittelung gewonnenen AEP-Zeitverlauf ausgemessen. Da es aber kaum noch Zweifel daran gibt, daß zur AEP-Generierung auch sinnesspezifische Hirnstrukturen wesentlich beitragen (Vaughan and Ritter, 1970; Kevanishvili et al., 1974; Perronet et al., 1974; Davis, 1976; Freigang et al., 1976), ist die Frage naheliegend, ob nicht der AEP-Zeitverlauf mehr diagnostisch nutzbare Information über den Funktionszustand des akustischen Sinneskanals enthält, als die übliche Ausmessung von nur 3 Werten liefert.

Zur Beantwortung dieser Frage muß mit geeigneten Verfahren geprüft werden, ob Erkrankungen des akustischen Sinneskanals zu spezifischen Veränderungen des AEP-Zeitverlaufs führen. Voraussetzung für eine diagnostische Nutzung ist dabei, daß sich eventuelle Unterschiede nicht nur als Mittelwertdifferenzen bei größeren Patientengruppen statistisch sichern lassen, sondern daß auch aus dem individuellen Potentialverlauf eines Patienten die Zugehörigkeit zur entsprechenden Patientengruppe mit ausreichender Sicherheit ableitbar ist. Nachfolgend wird über Ergebnisse von Untersuchungen hierzu berichtet.

## 2. Methodik

Die Bearbeitung der vorliegenden Aufgabenstellung erfordert mathematische Klassifikationsverfahren, die es ermöglichen,

<sup>1</sup>) HNO-Klinik des Bereichs Medizin (Charité) der Humboldt-Universität Berlin

<sup>2</sup>) Augenklinik des Bereichs Medizin (Charité) der Humboldt-Universität Berlin

<sup>3</sup>) ZKI der Akademie der Wissenschaften der DDR, Berlin



den gesamten AEP-Zeitverlauf in die Analyse einzubeziehen. Infrage kommen die multivariate Varianz- und Diskriminanzanalyse und Verfahren der automatischen Zeichenerkennung, mit beiden wurden erste Erfahrungen gesammelt. Ausgangswerte für diese Untersuchungen sind Trainingsstichproben mit Individuen bekannter Klassenzugehörigkeit (bekannte Diagnose), auf deren Grundlage das Prinzip der Klasseneinteilung ermittelt und daraus eine Regel zur richtigen Klassifizierung von Patienten mit unklarer Diagnose abgeleitet wird.

Um einen im Rahmen der angewendeten Klassifizierungsmodelle optimalen Klassifikator zu erhalten, müssen hinreichend große Stichproben vorliegen. Das Ziel der hier geschilderten Untersuchungen war jedoch noch nicht die Erstellung eines praktikablen Klassifikators, sondern es sollte zunächst geprüft werden, ob überhaupt der Zeitverlauf des AEP die vorgegebene Klasseneinteilung in Normalhörende und verschiedene otologische Patientengruppen widerspiegelt und in welcher Größenordnung der Klassifikationsfehler erwartet werden kann. Das ist schon mit Stichproben etwas geringeren Umfangs möglich. Die Anzahl der in die Untersuchungen einbezogenen Gruppen wurde auf folgende 3 Modellklassen beschränkt:

- |                              |                |
|------------------------------|----------------|
| 1. Normalhörende             | (76 Personen)  |
| 2. Patienten mit Lärmschaden | (35 Patienten) |
| 3. Patienten mit Hörsturz    | (16 Patienten) |

Da von vornherein keine Kenntnis darüber vorlag, bei welchen Reizparametern AEP erhalten werden, die besonders gut eine Klassentrennung ermöglichen (geringe Variabilität innerhalb der Klassen), wurden bei jeder Person mehrere AEP bei unterschiedlichen Reizbedingungen abgeleitet. Die zu variierenden Reizparameter waren die Reizstärke (wegen des in der Regel bei allen Patienten und Normalpersonen unterschiedlichen Hörverlustes erfolgten die Ableitungen bei gleichem »sensation level« (SL), d.h. bei gleichem Pegel über der individuellen Schwelle), das Interstimulus-Intervall (ISI) und die Reizfrequenz. Bei der Wahl von 2 Werten je Parameter ergeben sich 8 verschiedene Reizbedingungen für die AEP-Ableitung. Bei vielen Patienten erlaubte jedoch das Ausmaß der Schwellenerhöhung (vor allem bei 4000 Hz) keine Messungen bei 50 dB SL.

Zusätzlich zu diesen maximal 8 AEP je Person wurden durch Mittelwertbildung von 2 unter verschiedenen Reizbedingungen bei einem Patienten bzw. Normalhörenden abgeleiteten AEP sogenannte Kombinationspotentiale gewonnen:

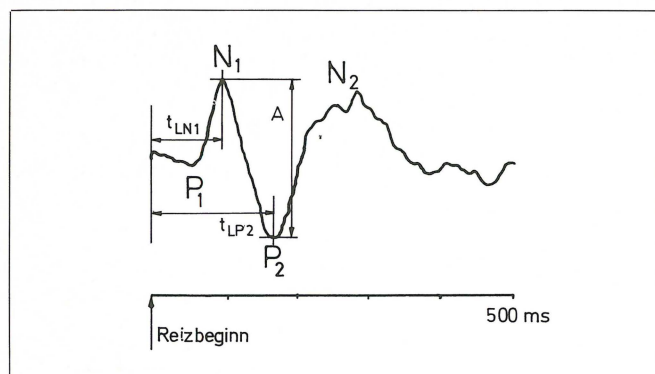
Kombinationspotential I:

Mittelwert der

AEP bei 1000 Hz, 30 dB SL, ISI = 2 s

und bei 4000 Hz, 30 dB SL, ISI = 2 s

Abb. 1. Übliche Ausmessung des AEP am Analogausschrieb (A = AEP-Amplitude,  $t_{LN1}$ ,  $t_{LP2}$  = Latenz von  $N_1$  bzw.  $P_2$ )



Kombinationspotential II:

Mittelwert der

AEP bei 4000 Hz, 30 dB SL, ISI = 2 s

und bei 4000 Hz, 30 dB SL, ISI = 5 s

Von diesen Kombinationspotentialen werden 2 Vorteile gegenüber den Einzelpotentialen erwartet:

1. Vergrößerung des Störabstands (die Mittelwertbildung ist bezüglich des Störabstands gleichbedeutend mit einer Erhöhung der Mittelungsschrittzahl von  $n = 64$  auf  $n = 128$ )
2. Eine über die Wirkung der Störabstandsverbesserung hinausgehende Erhöhung der Trennfähigkeit der durch die Kombination gewonnenen neuen Merkmale im Vergleich zu den Ausgangsmerkmalen.

Als Merkmale werden die den Potentialzeitverlauf charakterisierenden Stützstellen des AEP (Speicherung auf Lochstreifen) bezeichnet. Die Menge der  $p$  Merkmale bildet einen  $p$ -dimensionalen Merkmalsvektor ( $p = 100$ , bedingt durch den Averager DL 102 A). Vor der weiteren Analyse erfolgte eine Reduzierung der Merkmalsanzahl auf  $p = 50$  durch Weglassen jeder zweiten Stützstelle, um die Rechenzeiten des relativ langsamen LINC-8, auf dem die Diskriminanzanalyse gerechnet wurde, zu verringern. Das hat noch keinen Verlust relevanter Information zur Folge (Stürzbecher et al., 1978).

Zur Berücksichtigung der Altersabhängigkeit der AEP (Stürzbecher et al., 1977) wären für die Untersuchungen Stichproben mit stark eingeschränktem Altersbereich günstig. Wegen der relativ geringen Patientenzahlen mußte jedoch ein Bereich von 30–60 Jahren zugelassen werden, das mittlere Alter der Gruppen stimmt annähernd überein.

Eine ausführliche Beschreibung der verwendeten Klassifikationsverfahren würde in diesem Rahmen zu weit führen, deshalb werden beide nur soweit skizziert, wie es das Verständnis der Ergebnisse erfordert.

Multivariate Varianz- und Diskriminanzanalyse (Ahrens und Läter, 1974):

Vorausgesetzt wird, daß die Merkmale normalverteilten Grundgesamtheiten entstammen und die Kovarianzmatrizen der Klassen gleich sind. Die multivariate Varianzanalyse prüft, ob signifikante Unterschiede zwischen den Mittelwertvektoren der Klassen bestehen oder nicht. Außer einem summarischen Urteil über die Differenzen zwischen  $J$  Klassen ist auch ein Einzelvergleich von je 2 Klassen (Paarvergleich) möglich. Mittels der Diskriminanzanalyse wird auf der Grundlage der Trainingsstichproben der Klassifikator zur Trennung der Klassen berechnet. Dafür ist eine Reduzierung der vorgegebenen Merkmalsmenge durch Aussortierung der für die Trennung nicht relevanten Merkmale sinnvoll. Die Berechnung der nichtelementaren Diskriminanzmerkmale auf der Basis der reduzierten Merkmalsmenge ermöglicht dann eine Klassifizierung individueller AEP. Die erzielte Zuordnung kann durch ein zweidimensionales Diskriminationsschema graphisch veranschaulicht werden. In diesem Schema ist jeder Klasse ein kreisförmiger Streubereich zugeordnet, der die zur Klasse gehörenden Individuen mit der Wahrscheinlichkeit  $1-\alpha$  ( $\alpha = 0,05$ ) enthält.

Lernfähiger Trennebenenklassifikator (Unger et al., 1974; Unger et al., 1976):

Vor Erstellung des Klassifikators erfolgt eine Clusteranalyse aller Vorergebenen Klassen. Die Kenntnis von Objektballungen im Merkmalsraum innerhalb einer Klasse ermöglicht die Konstruktion eines stückweise linearen Klassifikators zur optimalen Trennung von nicht linear separierbaren Klassen. Dabei wird so vorgegangen, daß mittels der Methode der Regressionsanalyse (Meyer-Brötz und Schürmann, 1970) zwischen je zwei klassenfremden Clustern eine Trennebene erstellt wird. Das hierbei erhaltene Trennebenensystem ist nicht redundanzfrei, ersetzbare Ebenen müssen deshalb eliminiert werden. Die verbleibenden Ebenen bilden stückweise linear begrenzte Raumbereiche (Segmente), die den einzelnen Klassen zugeordnet werden. Ein Objekt  $x$  wird als zur Klasse  $C_i$  gehörend betrachtet, wenn es in einem Segment von  $C_i$  liegt. Liegt  $x$  in einem Gebiet, das zu keiner Klasse gehört, wird es der Klasse zugeordnet, zu deren Segment es den geringsten Abstand hat. In einer Belehrungsphase wird die Lage der Trennebenen, an denen Fehlentscheidungen auftreten, optimiert. Lernkriterium ist z.B. die Minimierung des gemittelten euklidischen Abstands der falsch klassifizierten Objekte zur Trennebene. Ein Vor-



teil des Verfahrens im Vergleich zur Diskriminanzanalyse ist die bessere Anpassung des Klassifikators bei von der Normalverteilung abweichenden Verteilungen bzw. beim Vorliegen unterschiedlicher Kovarianzmatrizen der Klassen.

Um die Realisierungsmöglichkeit für eine rechnergestützte Diagnosehilfe auf AEP-Grundlage einschätzen zu können, ist die Kenntnis des Klassifikationsfehlers wichtig, der beim praktischen Einsatz eines Klassifikators zu erwarten ist. Zur Schätzung dieses Fehlers sind verschiedene Verfahren bekannt (Zusammenfassung bei Toussaint and Sharpe, 1975), die drei wichtigsten werden nachfolgend kurz erläutert:

#### R-Methode

Nach Erstellung des Klassifikators auf der Grundlage der vorgegebenen Stichproben erfolgt eine Klassifizierung der Objekte dieser Trainingsstichproben. Der sich ergebende Prozentsatz falsch klassifizierter Objekte  $F[R]$  ist ein zu optimistischer Schätzwert für die wirkliche Fehlerrate, da sich der Klassifikator der in der Trainingsstichprobe realisierten zufälligen Objektkonfiguration anzupassen sucht.

#### U-Methode

Die U-Methode ermöglicht einen Test des Klassifikators mit für ihn fremden Objekten. Dazu wird jeweils ein Objekt der vorliegenden Stichproben ausgesondert, auf der Basis der übrigen  $n-1$  Objekte der Klassifikator bestimmt und damit dann das ausgesonderte Objekt klassifiziert. Diese Prozedur muß  $n$ -mal durchgeführt werden, bis alle Objekte einmal als Testobjekt klassifiziert worden sind. Die Differenz des so ermittelten Fehlerschätzwertes  $F[U]$  zu der bei einem praktischen Einsatz des Klassifikators zu erwartenden Fehlerrate ist bei angemessenem Stichprobenumfang nur gering, von großem Nachteil ist jedoch der erhebliche Rechenaufwand.

#### $\pi$ -Methode

Ein Kompromiß zwischen Rechenaufwand und Güte der Fehlerschätzung wird mit der  $\pi$ -Methode erreicht. Hierfür erfolgt nach Mischung mittels Zufallszahlen eine Einteilung der  $n$  Objekte in  $k$  Untermengen ( $\approx \frac{n}{k}$  Objekte je Untermenge). In  $k$  aufeinanderfolgenden Schritten wird jeweils eine Untermenge aussortiert, der Klassifikator auf der Grundlage der Objekte

der  $k-1$  Untermengen bestimmt und eine Klassifizierung der Objekte der ausgelassenen Untermenge durchgeführt. Die so ermittelte Fehlerrate  $F[\pi]$  ist ein zu pessimistischer Schätzwert für den zu erwartenden Fehler.

Wir führten Fehlerschätzungen sowohl nach der R-Methode als auch nach der  $\pi$ -Methode auf der Grundlage der vorliegenden kleinen Stichproben durch. Bei allen Fehlerschätzungen wurde nur die in den Merkmalen enthaltene Information über die Klassenzugehörigkeit berücksichtigt und auf die Ausnutzung von Informationen über die a priori-Wahrscheinlichkeit zur Resultatverbesserung bewußt verzichtet.

### 3. Ergebnisse

#### 3.1. Ergebnisse der einfachen Amplituden- und Latenz-ausmessung (univariate Analyse)

Bei allen abgeleiteten AEP erfolgte am Analogauschrieb eine Ausmessung der Amplituden und Latenzen (Abb. 1) und die Prüfung der Klassen-Mittelwertdifferenzen mit dem t-Test.

Da sowohl die Amplituden- als auch die Latenzmittelwerte der AEP der beiden pathologischen Gruppen bei keiner Reizbedingung signifikante Unterschiede aufweisen, ist eine Unterscheidung von Angehörigen dieser beiden Klassen auf der Basis eines einzelnen AEP-Merkmals nicht möglich.

Zwischen den Mittelwerten der Normalgruppe und beider pathologischen Gruppen lassen sich zwar für die  $N_1P_2$ -Amplitude und die  $N_1$ -Latenz bei 30 dB SL signifikante Differenzen (1 %-Niveau) nachweisen, eine Differenzierung zwischen normal und pathologisch ist jedoch auch hier mit Hilfe eines einzelnen Merkmals nicht mit ausreichender Sicherheit zu erzielen, da sich die Streubereiche der Klassen zu stark überlappen (Abb. 2).

#### 3.2. Ergebnisse der multivariaten Varianzanalyse

Das Ergebnis der Varianzanalyse der Einzel-AEP und der Kombinationspotentiale unter Einbeziehung aller 50 Merk-

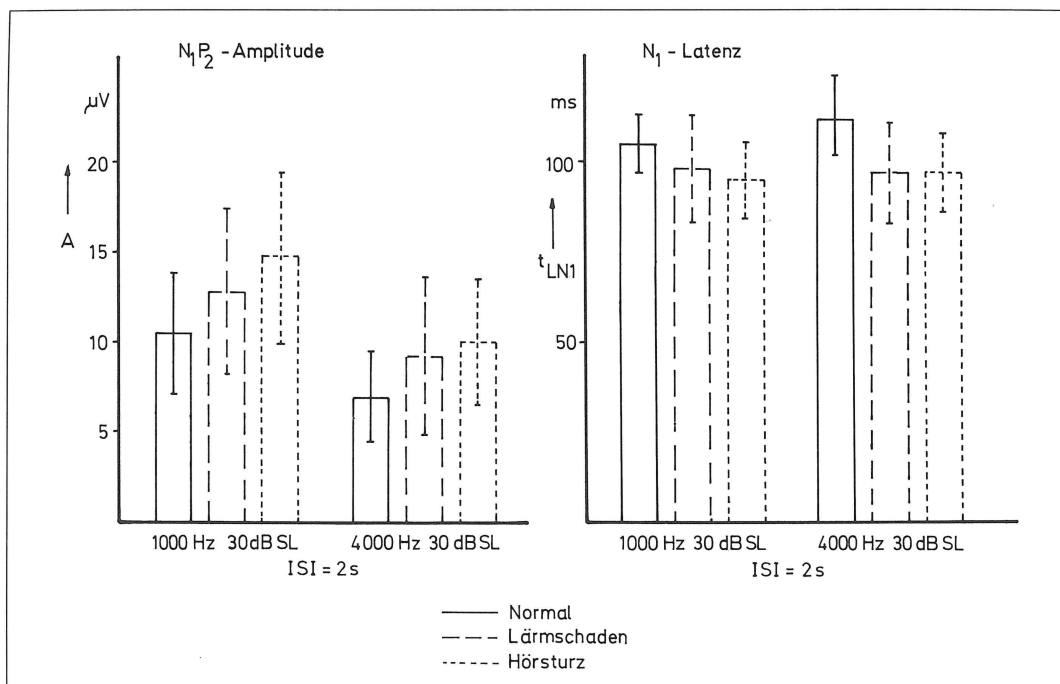


Abb. 2. Mittelwerte und Standardabweichungen der  $N_1P_2$ -Amplitude und der  $N_1$ -Latenz für die 3 Klassen bei 1000 Hz, 30 dB SL, ISI = 2 s.



Tab. 1. Ergebnis der Varianzanalysen der AEP auf der Grundlage von 50 Merkmalen. Es ist jeweils die Irrtumswahrscheinlichkeit angegeben, mit der die Nullhypothese der Gleichheit aller 3 Mittelwertvektoren (Gesamttest) bzw. der 2 Mittelwertvektoren des Paarvergleichs abgelehnt wird. Ein Strich bedeutet, daß die Nullhypothese nicht abgelehnt wird.

	Einzelpotentiale						Kombinationspotentiale	
	1000 Hz 50 dB SL ISI = 2 s	1000 Hz 30 dB SL ISI = 2 s	1000 Hz 50 dB SL ISI = 5 s	1000 Hz 30 dB SL ISI = 5 s	4000 Hz 30 dB SL ISI = 2 s	4000 Hz 30 dB SL ISI = 5 s	1000 Hz 4000 Hz 30 dB SL ISI = 2 s	4000 Hz 30 dB SL ISI = 2 s ISI = 5 s
Gesamt-F-Test	1 %	0,1 %	5 %	5 %	1 %	0,1 %	0,1 %	0,1 %
Paarvergleich $1/2$	5 %	1 %	—	—	1 %	0,1 %	0,5 %	0,1 %
Paarvergleich $1/3$	1 %	0,1 %	5 %	5 %	1 %	0,1 %	0,1 %	2,5 %
Paarvergleich $2/3$	—	—	—	—	—	0,5 %	0,1 %	—

male ist in Tabelle 1 dargestellt. Es ist jeweils die Irrtumswahrscheinlichkeit  $\alpha$  angegeben, mit der die Hypothese der Gleichheit aller 3 Mittelwertvektoren (Gesamttest) bzw. der Mittelwertvektoren des entsprechenden Paarvergleichs abgelehnt wird. Ein Strich bedeutet, daß die Nullhypothese nur bei einer Irrtumswahrscheinlichkeit von  $\alpha > 5\%$  abgelehnt wird.

Die statistisch am besten gesicherten Differenzen bei den Einzelpotentialen weisen die Klassenmittelwerte der bei 4000 Hz, 30 dB SL, ISI = 5 s abgeleiteten AEP auf. Alle Paarvergleiche sind hier hochsignifikant. Im Gegensatz dazu liegen bei den Reizparametern 1000 Hz, 50 dB SL, ISI = 5 s und 1000 Hz, 30 dB SL, ISI = 5 s nur statistisch nicht ausreichend gesicherte AEP-Mittelwertdifferenzen vor. Nur wenig besser sind die Ergebnisse bei 1000 Hz, 50 dB SL, ISI = 2 s. Die bei diesen 3 Reizbedingungen abgeleiteten AEP wurden daher nicht in die weiteren Analysen einbezogen.

Die Bildung des Kombinationspotentials I hat zu einer wesentlichen Erhöhung des Signifikanzniveaus der Paarvergleiche geführt. So ist der Paarvergleich  $2/3$  (Lärmpatienten/Hörsturzpatienten) bei beiden zugrunde liegenden Einzelpotentialen nicht, beim resultierenden Kombinationspotential I jedoch auf dem 0,1 %-Niveau signifikant. Es ist sehr unwahrscheinlich, daß eine derartige Erhöhung allein durch die relativ geringe Störabstandsverbesserung bewirkt wird, sondern man kann annehmen, daß durch die Kombination unabhängiger Merkmale ein zusätzlicher Gewinn erzielt worden ist.

Beim Kombinationspotential II wurde ein solcher Gewinn nur beim Paarvergleich  $1/2$  (Normal/Lärmschaden) erreicht. Das Signifikanzniveau dieses Vergleichs liegt höher als bei allen anderen Paarvergleichen  $1/2$ . Aus Tabelle 1 geht das nicht hervor, da der F-Wert nur bis 0,1 % tabelliert vorlag.

### 3.3. Ergebnisse der Diskriminanzanalyse

#### 3.3.1. Vergleich der verschiedenen Datensätze mittels des Fehlerschätzwertes der R-Methode

Anhand des mit der R-Methode erhaltenen Fehlerschätzwertes ist ein Vergleich der AEP bei den verschiedenen Reizbedingungen möglich. Dazu wurde bei jedem Datensatz eine Menge von 20 für die Trennung der Klassen relevanten Merkmalen aus den vorgegebenen 50 Merkmalen ausgewählt. Die Ergebnisse des Fehlertests sind in Tabelle 2 zusammengefaßt. Die Diskrimination der Einzelpotentiale ergibt im Vergleich mit den beiden Kombinationspotentialen ungünstigere Ergebnisse. Deshalb konzentrieren sich die weiteren Untersuchungen auf die letzteren. Für das Kombinationspotential I sind in

Abb. 3a die Mittelwertpotentiale der 3 Klassen dargestellt, Abb. 3b zeigt das Diskriminationsschema für die Objekte der Stichproben.

#### 3.3.2. Eingrenzung des für die Diskriminanzanalyse zu erwartenden Klassifikationsfehlers

Zur Eingrenzung des bei der Klassifizierung von Objekten unbekannter Klassenzugehörigkeit zu erwartenden Klassifikationsfehlers wurden für den Datensatz des Kombinationspotentials I die Fehlerraten nach der R-Methode und  $\pi$ -Methode in Abhängigkeit von der Merkmalsanzahl ermittelt. Für die Rechnungen mit der  $\pi$ -Methode erfolgte eine Einteilung der Gesamtmenge aller Individuen der vorliegenden 3 Stichproben nach Mischung mittels Zufallszahlen in 10 Untermenzen (10- $\pi$ -Methode). Das Ergebnis dieser Rechnungen ist in Abb. 4 dargestellt ( $\pi$ -Methode: oberste Kurve, gestrichelt; R-Methode: unterste Kurve, strichpunktiert).

Der Verlauf der mittels R-Methode erhaltenen Fehlerkurve zeigt, daß die Fehlerrate  $F[R]$  bei der Klassifizierung von Merkmalsvektoren, die bereits zur Klassifikatorerstellung beigetragen haben, eine monoton fallende Funktion der Merkmalsanzahl ist, da mit zunehmender Merkmalsanzahl die Verteilungsdetails der vorliegenden Stichproben immer besser erfaßt werden.

Auch die mit der  $\pi$ -Methode geschätzte Fehlerrate  $F[\pi]$  wird zunächst mit zunehmender Merkmalsanzahl kleiner, um dann jedoch nach Durchlaufen eines Minimums wieder anzusteigen. Ursache hierfür ist, daß die verhältnismäßig kleinen Stichproben keine ausreichend gute Schätzung der Verteilung in der Grundgesamtheit erlauben. Deshalb werden durch die weitere Erhöhung der Merkmalsanzahl zunehmend Verteilungszufälligkeiten der Stichproben erfaßt. Ein auf einer gro-

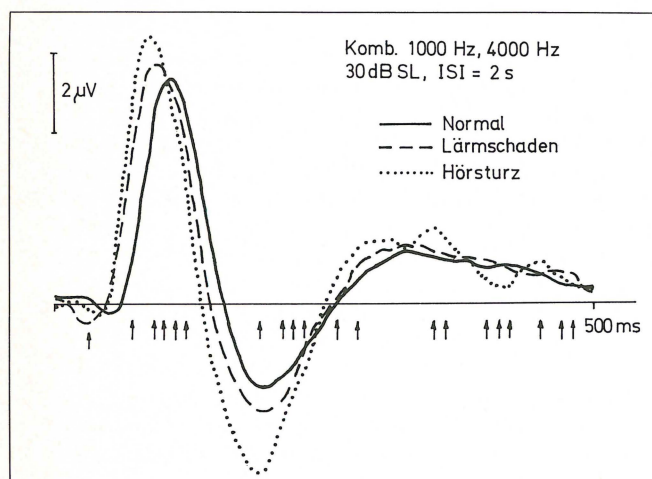
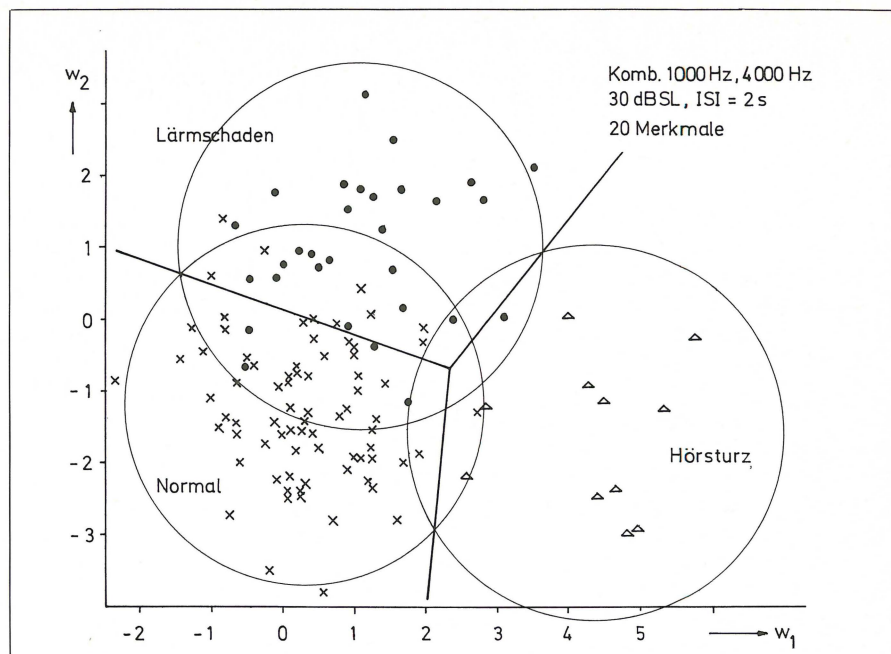
Tab. 2. Ergebnis der Fehlerschätzung mittels der R-Methode (20 Merkmale) für die AEP bei den verschiedenen Reizbedingungen.

	Einzel-AEP			Kombinations-AEP	
	1000 Hz 30 dB SL ISI = 2 s	4000 Hz 30 dB SL ISI = 2 s	4000 Hz 30 dB SL ISI = 5 s	I 1000 Hz 4000 Hz 30 dB SL ISI = 2 s	II 4000 Hz 30 dB SL ISI = 2 s ISI = 5 s
Fehler- rate	20,4 %	16,1 %	15,2 %	12,5 %	8,5 %



Abb. 3a (unten). Klassenmittelwerte des Kombinationspotentials I (1000 Hz/4000 Hz, 30 dB SL, ISI = 2 s). Die Pfeile geben die Lage der für die Diskriminanzanalyse (Abb. 3b) ausgewählten 20 Merkmale an.

Abb. 3b (rechts). Diskriminationsschema für das Kombinationspotential I auf der Basis der in Abb. 3a mit Pfeilen markierten 20 Merkmale. Die Kreise geben die 95 %-Streubereiche der Klassen an. Die Geraden ermöglichen eine Klassifizierung nach höchster Wahrscheinlichkeit.



Ben Merkmalsmenge basierender Klassifikator liefert dann bei Neuklassifikationen eine höhere Fehlerrate als ein mit wenigen Merkmalen erstellter. Dies wird auch von Trampisch (1975) und von Läuter (1976) betont.

Mit der R-Methode erhält man einen zu optimistischen Schätzwert für die zu erwartende Fehlerrate. Die  $\pi$ -Methode liefert dagegen einen zu pessimistischen Wert. Die Diskrepanz zwischen beiden Werten bzw. beider Werte zur wirklichen Fehlerrate ist auch bei kleinen Merkmalsmengen recht groß und erschwert die Einschätzung der Leistungsfähigkeit eines Klassifikators. Deshalb schlagen Toussaint und Sharpe (1975) vor, eine Gewichtsfunktion  $w$  so zu bestimmen, daß

$$w \cdot F[\pi] + (1 - w) \cdot F[R] \approx F[U]$$

ist, um damit bei Vermeidung des für die U-Methode erforderlichen großen Rechenaufwandes doch einen Fehlerschätzwert zu erhalten, der etwa dem der U-Methode entspricht und

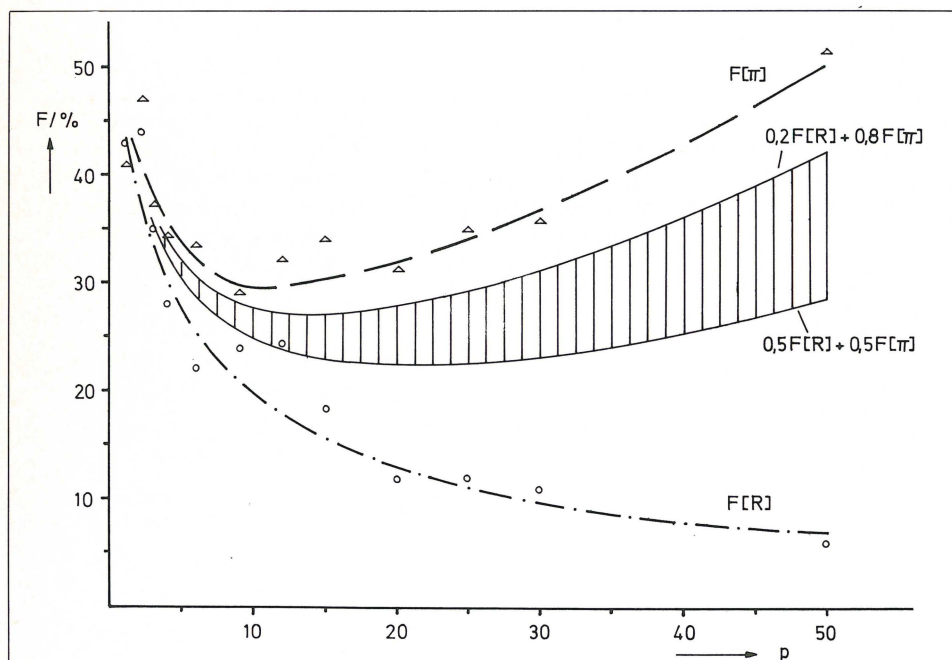


Abb. 4. Ergebnis der Fehlerschätzung für die Diskriminanzanalyse. Dargestellt sind die mittels R-Methode ( $F[R]$  – zu optimistische Schätzung) und  $\pi$ -Methode ( $F[\pi]$  – zu pessimistische Schätzung) in Abhängigkeit von der Merkmalsanzahl  $p$  berechneten Fehlerraten. Durch Wichtung beider Fehlerraten ( $w \cdot F[\pi] + (1 - w) \cdot F[R]$ ) läßt sich ein Bereich eingrenzen (schraffiertes Gebiet), in dem die Fehlklassifikationsrate des auf der Grundlage der vorliegenden kleinen Stichproben erstellten Klassifikators mit großer Wahrscheinlichkeit zu erwarten ist.



damit der wirklichen Fehlerrate nahekommte. Rein empirisch fanden die Autoren durch Berechnung der Fehlerschätzwerte nach der U-Methode, R-Methode und  $\pi$ -Methode (10- $\pi$ -Methode) auf der Grundlage einer 300 Patienten umfassenden Stichprobe, daß die Wahl von  $w = 0,5$  brauchbare Resultate liefert.

Auf Grund einer Abschätzung für multivariat normalverteilte Daten kommt McLachlan (1977) zu dem Schluß, daß die Wahl von  $w = 0,5$  nicht zulässig sei. Nach seinen Angaben muß der Schätzwert der R-Methode wesentlich geringer bewichtet werden als der der  $\pi$ -Methode, um eine gute Schätzung der zu erwartenden Fehlerrate zu erhalten.

In Abb. 4 wurden daher zusätzlich zu den Fehlerkurven der R- und  $\pi$ -Methode der arithmetische Mittelwert beider Kurven ( $w = 0,5$ ) sowie der Verlauf, der sich bei einer 0,2:0,8-Wichtung von R- und  $\pi$ -Methode ergibt, eingezeichnet. Innerhalb des von diesen beiden Kurven begrenzten Bereichs (schraffiert) liegt mit großer Wahrscheinlichkeit die Fehlerrate, die bei Neuklassifikationen mit einem auf der Grundlage der vorliegenden kleinen Stichproben erstellten Klassifikator zu erwarten ist. Mit 20 Merkmalen ist danach bei der Klassifizierung von Objekten unbekannter Klassenzugehörigkeit mit einer Fehlerrate zwischen 22 % und 27,5 % zu rechnen. Für einen in der Praxis einzusetzenden Klassifikator (große Trainingsstichproben) ist eine Verschiebung des angebbaren Fehlerbereichs zu geringeren Fehlerraten zu erwarten.

#### 3.4. Eingrenzung des Klassifikationsfehlers für den lernfähigen Trennebenenklassifikator

Die Fehlerschätzung erfolgte wie bei der Diskriminanzanalyse mittels R- und  $\pi$ -Methode (10- $\pi$ -Methode) in Abhängigkeit von der Merkmalsanzahl. Eine zusätzliche Clusterung der Objekte vor der Trennebenenstellung erwies sich als nicht sinnvoll, da der Umfang der Stichproben dafür zu gering ist. Die Ergebnisse der Fehlerschätzung sind in Abb. 5 dargestellt.

Eine wesentliche Differenz zu den mit der Diskriminanzanalyse gewonnenen Ergebnissen zeigt sich nur bei der Fehlerschätzung mit der R-Methode bei einer Merkmalsanzahl  $p > 20$ . Dieser Unterschied war zu erwarten, da infolge der Einzelherstellung der Ebenen und der Belehrung mit den falsch klas-

sifizierten Objekten die Verteilungsdetails der Stichproben besser erfaßt werden. Die Folge ist eine kleinere Fehlerrate des Trennebenenklassifikators beim Fehlertest mit der R-Methode.

Dagegen unterscheiden sich die Ergebnisse der  $\pi$ -Methode bei der Diskriminanzanalyse und dem lernfähigen Trennebenenklassifikator nur unwesentlich. Ein besseres Ergebnis mit dem Trennebenenklassifikator wäre dann zu erwarten (bei ausreichend großen Stichproben), wenn wesentliche Abweichungen von der mehrdimensionalen Normalverteilung vorliegen und die Kovarianzmatrizen der Klassen große Unterschiede aufweisen. Die annähernde Übereinstimmung der  $\pi$ -Methoden-Schätzwerte bei Diskriminanzanalyse und lernfähigem Trennebenenklassifikator kann daher eventuell als Hinweis auf das Vorliegen einer Normalverteilung bei den der Untersuchung zugrunde liegenden Daten gewertet werden.

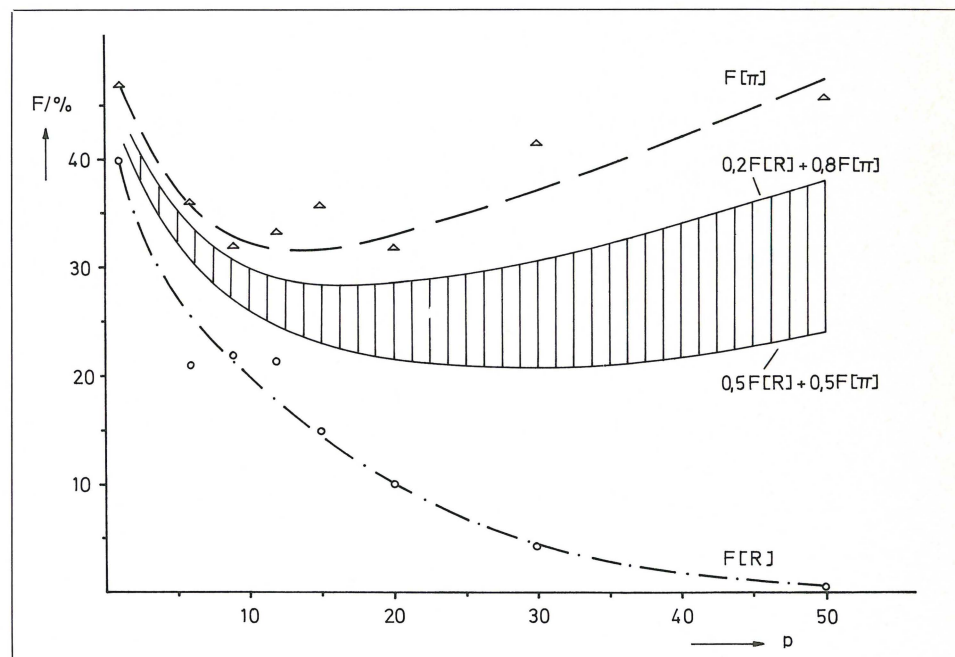
Das schraffierte Gebiet in Abb. 4 stellt wie für die Diskriminanzanalyse in Abb. 4 den Bereich dar, in dem die Fehlerrate des auf der Grundlage der vorhandenen kleinen Stichproben erstellten lernfähigen Trennebenenklassifikators zu erwarten ist. Die Diskussion im Abschnitt 3.3.2. zur weiteren Eingrenzung des Fehlers bei der Diskriminanzanalyse gilt ganz analog auch für den lernfähigen Trennebenenklassifikator und wird deshalb hier nicht wiederholt.

#### 4. Schlußfolgerungen für die praktische Nutzung

Die durchgeführten Untersuchungen zur Eingrenzung des Klassifikationsfehlers haben ergeben, daß der Fehler eines auf der Basis der vorliegenden kleinen Stichproben erstellten Klassifikators zwischen 22 % und 27 % liegt. Für einen Klassifikator, dem wesentlich größere Patientenstichproben zugrunde liegen, ist mit einer kleineren Fehlerrate zu rechnen.

Der zu erwartende Fehler ermöglicht zwar, wie vorauszu-  
sehen war, keine ausschließlich auf dem AEP basierende otologische Diagnostik, mit dem AEP-Klassifizierungsergebnis wird dem Arzt jedoch eine wichtige zusätzliche Information zur Verfügung gestellt, die die Ergebnisse herkömmlicher diagnostischer Verfahren ergänzt und damit eine sicherere Diagnosestellung ermöglicht (rechnerunterstützte Diagnostik).

Abb. 5. Ergebnis der Fehlerschätzung für den lernfähigen Trennebenenklassifikator. Dargestellt sind die mittels R-Methode ( $F[R]$  – zu optimistische Schätzung) und  $\pi$ -Methode ( $F[\pi]$  – zu pessimistische Schätzung) in Abhängigkeit von der Merkmalsanzahl  $p$  berechneten Fehlerraten. Durch Wichtung beider Fehler ( $w \cdot F[\pi] + (1 - w) \cdot F[R]$ ) läßt sich ein Bereich eingrenzen (schraffiertes Gebiet), in dem die wirkliche Fehlerrate des auf der Grundlage der vorliegenden kleinen Stichproben erstellten Klassifikators mit großer Wahrscheinlichkeit zu erwarten ist.





Für den praktischen Einsatz des AEP-Klassifikators ist es erforderlich, alle Möglichkeiten zur Reduzierung der Fehler-rate auszuschöpfen. Die bereits erwähnte Vergrößerung der Trainingsstichproben ist hier von wesentlicher Bedeutung. (Das betrifft sowohl die schon vorliegenden als auch die weiteren entsprechend den medizinischen Fragestellungen noch zu erhebenden Stichproben.)

Bei der Datenerfassung wird es sicher sinnvoll sein, von bisher 64 Mittelungsschritten je AEP auf 128 überzugehen, da hiermit sowohl eine Störabstandsverbesserung als auch eine Verringerung der intraindividuellen Kurzzeit-Variabilität des AEP zu erzielen ist.

Auf der Verarbeitungsseite ist die Realisierung des von Läuter (1976) vorgeschlagenen Verfahrens zur kombinierten Anwendung mehrerer Diskriminationsmodelle zu empfehlen, da nach den Angaben des Autors damit ein kleinerer Fehler als bei der üblichen Anwendung einer einzelnen Diskriminationsregel zu erzielen ist.

Vorteile bei der Anwendung des lernfähigen Trennebenenklassifikators gegenüber der Diskriminanzanalyse konnten zwar für die vorliegenden Daten nicht nachgewiesen werden, für große Stichproben müßte diese Frage jedoch nochmals überprüft werden.

## Literatur

- AHRENS, H., LÄUTER, J. (1974): Mehrdimensionale Varianzanalyse. Akademie Verlag Berlin.
- DAVIS, H. (1976): Principles of electric response audiometry. *Ann. Otol. Rhinol. Laryngol.* 85 Suppl. 28.
- FREIGANG, B., VON SPECHT, H., KEVANISHVILI, Z. S. (1976): Experimentelle Hinweise auf den unterschiedlichen Entstehungsmechanismus des späten akustisch evozierten Potentials. Vortrag auf dem II. Bilateralen ERA-Symposium, Eisenach.
- KEVANISHVILI, Z. S., VON SPECHT, H., KAJAJA, O. A. (1974): Amplitude and latency studies of the human slow auditory evoked responses to tones of different frequencies. *Estrat. da Minerva Otorinolaringologica* 24, 230.
- LÄUTER, J. (1976): Kombinierte Anwendung mehrerer Diskriminationsmodelle zur Erniedrigung des Zuordnungsfehlers. Vortrag auf der Tagung des Arbeitskreises Mathematische Kybernetik, Berlin.
- MCLACHLAN, G. J. (1977): A note on the choice of a weighting function to give an efficient method for estimating the probability of misclassification. *Pattern Recognition* 9, 147.
- MEYER-BRÖTZ, G., SCHÜRMANN, J. (1970): Methoden der automatischen Zeichenerkennung. Akademie Verlag Berlin.
- PERRONET, P., MICHEL, F., ECHALLIER, J. F., GIROD, J. (1974): Coronal topography of human auditory evoked responses. *Electroenceph. clin. Neurophysiol.* 37, 225.
- STÜRZEBECHER, E., WERNECKE, K.-D., WAGNER, H., GERHARDT, H.-J. (1977): Der Nachweis von Altersveränderungen des langsamen akustisch evozierten Potentials bei Erwachsenen mittels der multivariaten statistischen Analyse. *Arch. Oto-Rhino-Laryng.* 218, 19.
- STÜRZEBECHER, E., WERNECKE, K.-D., WAGNER, H., GERHARDT, H.-J. (1978): Multivariate Untersuchungen zur intra- und interindividuellen Variabilität der langsamen akustisch evozierten Potentiale (AEP). *Laryng. Rhinol.* 57, 63.
- TOUSSAINT, G. T., SHARPE, P. M. (1975): An efficient method for estimating the probability of misclassification applied to a problem in medical diagnosis. *Comput. Biol. Med.* 4, 269.
- TRAMPISCH, H. J. (1975): Trennprobleme bei unvollständiger Information – eine Übersicht. *EDV in Medizin und Biologie* 6, 2.
- UNGER, S., JENTSCH, W., STANKE, G. (1974): Stückweise lineare Klassifizierung durch Maximierung eines statistischen Gütefunktionalen über einem automatisch erzeugten minimalen System von Trennebenen bei adaptiver Entscheidungslogik. Forschungsbericht des ZKI der AdW der DDR.
- UNGER, S., JENTSCH, W., TÜRSCHMANN, M. (1976): Die mit Trennebenen arbeitenden Rechenprogramme eines komplexen Klassifizierungssystems. *ZKI Informationen – Lernende Systeme – 3*, AdW der DDR.
- VAUGHAN, H. G., RITTER, W. (1970): The sources of auditory evoked responses recorded from the human scalp. *Electroenceph. clin. Neurophysiol.* 28, 360.

Anschrift der Autoren: Dipl.-Ing. E. Stürzebecher, Dr.-Ing. H. Wagner und Prof. Dr. sc. med. H.-J. Gerhardt, HNO-Klinik des Bereichs Medizin (Charité) der Humboldt-Universität Berlin, Schumannstr. 20, DDR-104 Berlin; Dr. rer. nat. K.-D. Wernecke, Augenklinik des Bereichs Medizin (Charité) der Humboldt-Universität Berlin, Ziegelstr., DDR-104 Berlin; Dr. rer. nat. S. Unger, Zentralinstitut für Kybernetik und Informationsprozesse der Akademie der Wissenschaften der DDR, Rudower Chaussee, DDR-1199 Berlin-Adlershof.



# Neue Möglichkeiten für nichtsehende Mitarbeiter in der Datenverarbeitung

## Situationsanalyse und Darstellung eines Forschungsvorhabens

E.-D. Lorenz

### Zusammenfassung

Die von Rechnern ausgegebenen Informationen sind nahezu ausschließlich für die Aufnahme durch das Auge bestimmt. Nichtsehende Mitarbeiter in der Datenverarbeitung sind daher auf die Unterstützung durch einen Mitarbeiter angewiesen. Für die Umsetzung der Rechnerein- und -ausgaben in eine Blinden zugängliche Darstellungsweise bietet sich bei dem heutigen Stand der Technik fast ausschließlich die Braille'sche Punktschrift an. Das Angebot an derartigen DV-Endgeräten ist bisher sehr gering. Drei Brailleschriftgeräteformen scheinen soweit ausgereift zu sein, daß sie im Zusammenwirken mit DV-Anlagen einsetzbar bzw. adaptierbar sind: Brailleschrift-Streifenschreiber, Blattschreiber und Braille-Zeilengeräte. Diese sollen in einem von der DFG geförderten Forschungsvorhaben auf ihre spezifischen Einsatzmöglichkeiten hin untersucht werden. Dafür wurde das Sechs-Punkte-Brailleschrift-System zu einem Acht-Punkte-Code erweitert. Neben dem lokalen Einsatz der Geräte sollen auch Fragen des Zusammenwirkens mit Großrechnern erörtert werden. Die Unabhängigkeit des Nichtsehenden kann jedoch nur dann erreicht werden, wenn auch die Software an seine spezifischen Belange angepaßt ist. Hierfür und zur Steigerung der Portabilität des angestrebten Systems sind Softwarearbeiten geplant.

### Summary

The information put out by computers are nearly exclusively determined for the exposure through the eye. For this reason blind people in data processing depend on the assistance of a collaborator. For conversion of input and output, readable for the visually handicapped people, up to now only the Braille characters (typing) are available. The offer of I/O devices is very small. It seems that three types of Braille units in development matured out so that they can act with computer I/O devices: Braille-paper tape embosser, page embosser and Braille-line embosser. These units should be examined in practical applications in a research project by the Deutsche Forschungsgemeinschaft. Therefore the Braille six-point-printer-system is extended to an eight-point-code. Beside the local operation interfacing and communication shall be investigated. The independence of visually handicapped people can only be reached, if additionally software is adapted on their specific importance. For this reason as well as for the portability of the system to be developed, software improvements are planned.

### 1. Einleitung

Fast alle Informationen, die von Rechenanlagen ausgegeben werden, sind für die Aufnahme durch das Auge bestimmt. Daher sind die Arbeitsmöglichkeiten für Nichtsehende in der Datenverarbeitung stark eingeschränkt, weil sie auf ständige Unterstützung durch einen Mitarbeiter (Vorleser) angewiesen sind.

Bei der Kommunikation zwischen nichtsehenden DV-Fachleuten und Rechnern ist die Informationseingabe im wesentlichen problemlos. Sie kann über die Tastatur der Datenendgeräte erfolgen, die mit speziellen Tastmarkierungen versehen sein sollten. Allerdings ist für den Nichtsehenden die Möglichkeit des Nachlesens der geschriebenen Texte nicht gegeben, so daß die vom Endgerät bereitgestellten Korrekturmöglichkeiten nur teilweise ausgenutzt werden können.

Die hauptsächliche Schwierigkeit liegt in der Verarbeitung von Rechnerausgaben, d.h. in der Erkennung, Selektion und Speicherung der über Datenendgeräte schreibend, anzeigend oder druckend übermittelten Antworten und Ergebnisse.

Für die Umsetzung der Rechnerein- und -ausgaben in eine dem Nichtsehenden zugängliche Darstellungsweise bietet sich bei dem heutigen Stand der Technik fast ausschließlich die Braille'sche Punktschrift im Sechs-Punkt- und versuchsweise im Acht-Punkt-Code an. Bei diesem von dem Franzosen Louis Braille 1825 entwickelten System werden die Zeichen (Buchstaben, Ziffern, Sonderzeichen) aus den sechs Punkten einer  $3 \times 2$  Matrix gebildet. Für die Bedeutung eines Zeichens sind Anzahl und Stellung der Punkte innerhalb der Grundform und bei komplizierteren Texten, wie Kurzschrift, mathematischen Formeln oder Noten, auch die Stellung der Zeichen innerhalb des Kontextes entscheidend. Aus den sechs Punkten können einschließlich Leerfelder 64 Zeichen gebildet werden.

### 2. DV-Endgeräte

Bedingt durch die zahlenmäßig sehr geringe Nachfrage sind DV-Endgeräte für Blinde bisher kaum erhältlich. Drei Brailleschriftgeräteformen scheinen mir indessen soweit ausgereift zu sein, daß sie als Datenendgeräte oder Zusatzgeräte zu Endgeräten einsetzbar bzw. adaptierbar sind.

Dies sind:

#### 1. Brailleschrift-Streifenschreiber

Ein solches Gerät wurde auf meine Anregung hin von der



Firma Thiel in Jugenheim zum Anschluß an ein Datenendgerät weiterentwickelt. Es ist seit Herbst 1975 zu einem Terminal IBM 2741 der Medizinischen Hochschule Hannover parallel geschaltet (s. Abb. 1). Die über dieses Terminal laufenden Ein- und Ausgaben können damit unmittelbar in Punkschrift übertragen werden.

Hierfür mußte das 64 Zeichenmöglichkeiten bietende Sechs-Punkte-Brailleschrift-System ( $3 \times 2$  Matrix) zu einem Acht-Punkt-System ( $4 \times 2$  Matrix) erweitert werden, da mit Steuersignalen 93 Zeichen übertragen werden müssen. Mit diesem Gerät ist eine Online-Informationserkennung über einige Zeilen möglich. Die serielle Ausgabe auf Endlosband erschwert jedoch die Auswertung und insbesondere die Speicherung größerer Rechnerausgaben. Das Gerät ist daher vorwiegend für interaktiven Betrieb, wie Editing-Arbeiten und die Verwendung interaktiver Programmsysteme mit geringem Ausgabeumfang geeignet. Vorteilhaft ist, daß das Terminal durch die charakteristische Rhythmik des Schreibkopfes (Leerzeichen, Zeilenlänge, Zeilenvorschübe) zusätzliche akustische Informationen liefert, die dem Nichtsehenden das Verfolgen bekannter Sequenzen erleichtern.

## 2. Blattschreiber

Mit einem solchen Gerät können der Druckerausgabe entsprechende Brailleschrifttexte erzeugt werden. Allerdings sind die angebotenen Geräte verhältnismäßig langsam. Außerdem ist das Volumen eines Brailleschrifttextes ungefähr 50mal so groß wie das einer Normalschriftvorlage. Dieses enorm ungünstige Verhältnis ist durch die für die Ertastung erforderliche Zeichengröße von etwa einem qcm und die für die Haltbarkeit der Punkte nötigen Papierdicke bedingt. Solche Geräte dürften daher für die Ausgabe wesentlicher Informationen möglichst begrenzten Umfangs gut einsetzbar sein. Es müssen also Selektionsverfahren geschaffen werden, um die Ausgabeumfänge zu reduzieren. Dafür müssen die hard- und softwaremäßigen Voraussetzungen gegeben sein bzw. geschaffen werden. Die angebotenen Geräte haben nur eine Zeilenlänge von max. 40 Zeichen. In der Datenverarbeitung sind jedoch 80 Zeichen (Lochkarte, Bildschirm) oder 133 bzw. 136 Zeichen bei den meisten Schnelldruckern gebräuchlich. Hierdurch ergeben sich Probleme beim Übertragen und Auswerten von Tabellen, die beim Einsatz eines Brailleschrift-Blattschreibers zusätzliche Softwarearbeiten erfordern.

Abb. 1. Terminal IBM 2741 mit zugeordnetem Acht-Punkte-Brailleschrift-Streifenschreiber rechts

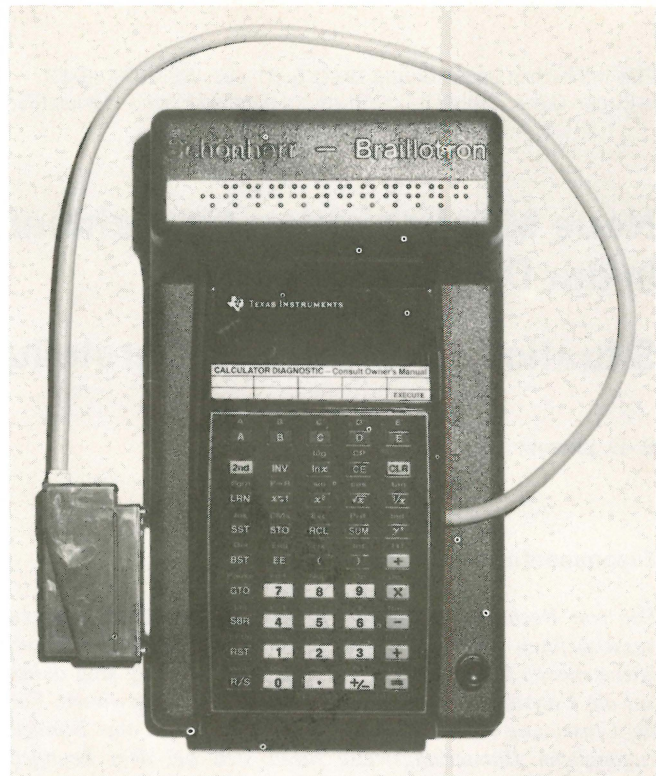
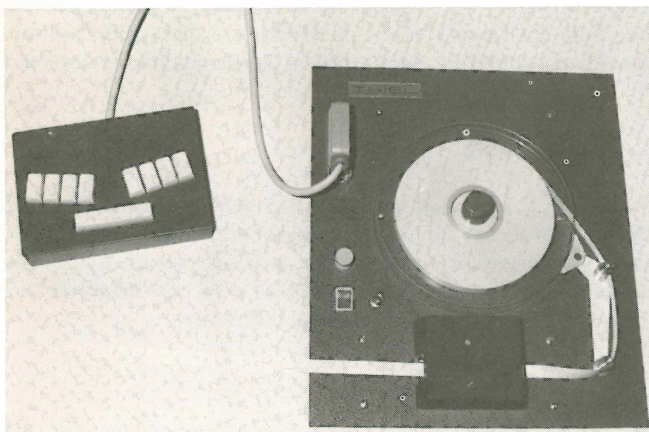


Abb. 2. Taschenrechner mit Braillezeilenzusatzgerät

## 3. Braille-Zeilengeräte

Die Braillezeichen erscheinen hierbei nicht auf Papier sondern werden durch elektromagnetisch angesteuerte Metallstifte erzeugt, sind also wieder löschar bzw. wieder neu setzbar wie beim Bildschirm. Diese Braillemodultechnik ist im Rahmen eines vom BMFT geförderten Forschungsprojektes entwickelt und zunächst für den begrenzten Zeichenvorrat von Taschenrechnern praktisch eingesetzt worden (s. Abb. 2).

Entwickelt wurde in dieser Technik:

- Ein Zusatzgerät zum Bildschirm, mit dem jeweils eine über eine Spezialtastatur anwählbare Bildschirmzeile von 80 Zeichen Länge in zwei übereinander angeordneten Halbzeilen tastbar gemacht werden kann.
- Ein eigenständiges Brailleterminal mit Eingabetastatur und Ausgabemöglichkeit für eine Braillezeile von 48 Zeichen.

Braille-Zeilengeräte werden sich voraussichtlich für den interaktiven Betrieb und insbesondere zur schnellen Erfassung und Selektion größerer Informationsmengen eignen, da sie in gewisser Weise einen direkten Zugriff auf Arbeits- und Masenspeicher erlauben.

Die drei beschriebenen Gerätetypen haben spezifische Vorteile aber auch Beschränkungen. In einer Pilotstudie sollen daher die neuen Möglichkeiten erforscht werden, die für Nichtsehende durch den Einsatz der genannten Geräte oder Gerätekombinationen geschaffen werden können. Dazu müssen die drei Brailleschriftgeräte auf ihre spezifischen Einsatzmöglichkeiten hin untersucht werden.

## 3. Das DFG Forschungsvorhaben

Die Aufgabenstellung eines von der DFG finanzierten Forschungsvorhabens »Untersuchungen zu Brailleschriftum-



kodierung, Brailleendgeräteinsatz und Softwareentwicklung für blinde DV-Fachleute« besteht darin, Wege aufzuzeigen, wie sich für Nichtsehende Möglichkeiten zu selbständigem Arbeiten eröffnen lassen.

Eine wesentliche Grundlage hierfür ist die Erweiterung und Umkodierung des Brailleschrift-Systems zu einem Acht-Punkt-Code. Hierzu wurden den 128 Zeichen des 7-Bit-ASCII-Codes geeignete Zeichen der Acht-Punkt-Schrift zugeordnet. Dabei wurde Wert darauf gelegt, den 64-Zeichen-ASCII-Satz auf die aus sechs Punkten gebildeten Zeichen der Braille'schen Grundform abzubilden. Diese Zuordnung erlaubt es bei einer Reihe von Anwendungen mit Sechspunkt-Zeichen auszukommen. In diesen Fällen kann damit die durch die unteren beiden Ergänzungspunkte entstehende Vergrößerung der Brailleschriftausgabe um weitere 25 % vermieden werden. Eine gewisse Schwierigkeit könnte sich allerdings beim Ertasten von aus nur einem Punkt gebildeten Zeichen innerhalb des Acht-Punkt-Systems ergeben. Es muß in der Versuchsphase noch erprobt werden, ob sie jeweils eindeutig und gut erkennbar sind. Eine Änderung der den Braille-Code generierenden PROMs ist technisch leicht möglich.

### 3.1 Hardware

Für dieses Forschungsvorhaben hat die DFG als lokale Gerätekonfiguration zur Verfügung gestellt:

1. als Zentraleinheit den Kleinrechner MODCOMP II/12 mit 64 kB Arbeitsspeicher und  $2 \times 2,5$  MB Doppelwechselplattensystem für lokale Datenhaltung; die Software umfaßt ein komfortables Betriebssystem mit Online-Editor und File Manager, einen leistungsfähigen FORTRAN-Compiler sowie einen UT200-Emulator zum Anschluß an die CYBER 73/76 des Regionalen Rechenzentrums Niedersachsen.
2. ein »intelligentes« Datensichtgerät (Beehive 500).
3. ein Braillezeilenzusatzgerät der Firma Schönherr in Stuttgart zur Übertragung einer anwählbaren Bildschirmzeile in Punktschrift.
4. das Brailleschrift-Blattschreiberterminal LED-120 der Firma Triformation in Stuart, Florida, USA.

Der Preis für das Gesamtsystem beläuft sich auf weniger als 150 000 DM.

Der MODCOMP-Rechner dient als Front-End-Prozessor für die CYBER 73/76 des Regionalen Rechenzentrums Niedersachsen, an die er mittels UT200-Emulator für synchrone Datenübertragung angeschlossen ist. Darüber hinaus werden folgende Verbindungen zur IBM 360/67 der Medizinischen Hochschule Hannover angestrebt:

1. Anschluß des vorhandenen Terminals IBM 2741 an die MODCOMP,
2. Verbindung der MODCOMP zum interaktiven CMS und
3. Anschluß der MODCOMP über BSC-2780 Synchronprozedur an das OS.

### 3.2 Software

Mit der Installation der beschriebenen Konfiguration und Realisierung der genannten Anschlüsse sind die Voraussetzungen dafür geschaffen, Softwarearbeiten durchzuführen, die dem Nichtsehenden selbständiges Arbeiten ermöglichen sollen. Dazu muß den spezifischen Belangen des Blinden und den Spezialgeräten angepaßte Software erstellt werden. Im Vordergrund des Forschungsvorhabens stehen daher folgende Punkte:

- Adaption des Betriebssystems zur effektiven Kombination der Brailleschrift-DV-Endgeräte,
- Behandlung und Umsetzung von Interrupts in eine von Nichtsehenden erkennbare und interpretierbare Form (eventuell auch akustisch),
- Erstellung von Selektionsroutinen zur Reduktion des Ausgabeumfanges auf ein für den Brailledrucker sinnvolles Maß,
- Untersuchung der Umformatierungsmöglichkeiten von Ausgabebetabellen auf die nur 40 Zeichen lange Zeile des Brailledruckers.

In diesem Zusammenhang ist zu überprüfen, ob die Ausgabebetabellen der für statistische Auswertungen klinischer Studien geeigneten Programmsysteme wie SPSS und BMD diesem Format angepaßt werden können.

Um die Portabilität des angestrebten Systems zu erhöhen, ist schließlich zu erörtern, ob und ggf. wie die Anpassungsprobleme der Geräte untereinander und im Zusammenwirken mit Großrechnern generell beschrieben werden können. Hier sind Fragen der geeigneten Übertragungsgeschwindigkeit, des Übertragungsmodus, sowie der Codegenerierung und -anpassung zu behandeln.

## 4. Schlußbemerkungen

Das Forschungsvorhaben soll zeigen, daß mit kostenmäßig und technisch vertretbarem Aufwand ein von Nichtsehenden selbständig bedienbares Interaktionssystem geschaffen werden kann. Es ist daher zu erwarten, daß ihnen dadurch neue Arbeitstechniken und damit neue Betätigungsmöglichkeiten in der Datenverarbeitung eröffnet werden können.

Die angestrebten Fortschritte in Richtung auf dieses Ziel bilden somit gleichzeitig einen Beitrag zur Rehabilitation Nichtsehender.



## Auswertung von Daten aus dem Pankreasfunktionstest mit drei multivariaten Methoden: Clusteranalyse, nichtlineare Diskriminanzanalyse und Faktorenanalyse

J. B. Schäffer, R. Thurmayr, G. R. Thurmayr, M. Otte

### Zusammenfassung

*Der Pankreasfunktionstest lieferte die Daten, auf die drei multivariate Methoden – Clusteranalyse, nichtlineare Diskriminanzanalyse und Faktorenanalyse – angewendet wurden. Nach einer Datenreduktion wurde die endgültige Auswertung mit 13 numerischen Variablen von 242 Personen vorgenommen, von denen die Außenkriterien, nämlich die klinischen Diagnosen, bekannt waren.*

*Die Ergebnisse wurden mittels Matrizen, Clusterdiagrammen und Dendrogramm dargestellt, wobei man findet, daß die Diskriminanzanalyse bessere Trefferraten als die Clusteranalyse liefert, wohingegen letztere die zugrundeliegende Datenstruktur sehr instruktiv aufzeigt. Mit der Faktorenanalyse wird eine Interpretation der wichtigsten Faktoren versucht.*

### Summary

*The pancreatic function test yielded the data to which three multivariate methods – cluster analysis, nonlinear discriminant analysis and factor analysis – were applied. After a data reduction the final evaluation was carried out with data of 13 numerical variables of 242 persons whose outer criteria – the clinical diagnoses – were known.*

*The results presented by means of matrices, clusterdiagrams and dendograms show clearly that discriminant analysis yields a better accuracy rate, whereas cluster analysis shows instructively the underlying structure of the data. With factor analysis the principal factors are tried to be interpreted.*

Der Pankreasfunktionstest wird in der Diagnostik von Bauchspeicheldrüsenerkrankungen eingesetzt, um Funktionseinschränkungen des Pankreas infolge einer Erkrankung erkennen zu können. Durch chemische Analyse des Pankreassaftes werden eine Reihe von Meßwerten gewonnen, die dem Arzt eine Beurteilung der Pankreasfunktion gestatten. Man unterscheidet danach eine pathologische, verminderte und normale Pankreasfunktion. Der Geübte erkennt die Extremfälle leicht; im Graubereich der mäßigen Funktionseinschränkung ist dagegen die Diagnosenstellung unmöglich. Aus diesem Grunde hat man für die Diagnostik eine Unterstützung durch Rechenverfahren gesucht.

### Pankreasfunktionstest

Der Pankreassaft wird durch eine Sonde gewonnen, die im Zwölffingerdarm bis zur Mündungsstelle des Bauchspeicheldrüsenganges vorgeschoben wird. Hier wird eine Saftmischung abgesaugt, die vorwiegend Pankreassaft enthält, jedoch auch

mit Galle aus dem ebenfalls hier mündenden Gallengang und Duodenalsaft angereichert ist. Um die Beimischung von Magensaft zu verhindern, wird eine doppeläufige Sonde eingelegt, die am Magenausgang den Magensaft absaugt.

Die Bauchspeicheldrüse wird während des Testes durch die Injektion von Sekretin und Cholecystokinin zur Sekretion angeregt. Eine 1. Injektion verursacht einen Sekretionsstoß in etwa physiologischem Ausmaß und eine 2. Injektion in einstündigem Abstand eine Maximalreizung der Bauchspeicheldrüse. In beiden Reizphasen wird das Pankreassekret in je 3 Fraktionen von 20 Minuten Dauer gesammelt. Gleichzeitig wird zu 3 Zeitpunkten eine Blutprobe abgenommen.

Innerhalb der Saftfraktionen werden jeweils die Saftmenge und der pH-Wert bestimmt und durch chemische Analyse die Konzentration der Pankreas-Enzyme Amylase, Lipase, Trypsin und die von Bilirubin und von Bicarbonat gemessen. Die ausgeschiedene Menge der Saftanteile wird aus dem Produkt von Konzentration und Saftvolumen berechnet. Im abgenommenen Blut wird die Serumlipase bestimmt, die einen Hinweis auf die Stärke des Übertrittes von Lipase aus der Bauchspeicheldrüse in das Blut gibt. Als Endergebnis der chemischen Analyse liegen dem Arzt 77 Meßwerte zur Beurteilung vor.

### Dokumentation des Pankreasfunktionstests

Um ein Rechenverfahren für den Pankreasfunktionstest erarbeiten zu können, wurde ein Dokumentationsverfahren aufgebaut, mit dessen Hilfe neben den Meßwerten aus dem Pankreasfunktionstest, Laborwerte, anamnestiche Angaben und klinische Befunde erfaßt und gespeichert wurden. Als Vergleichsdaten wurden die subjektive Testbeurteilung und als Außenkriterien die endgültigen Diagnosen mit dokumentiert, die histologisch oder endoskopisch gesichert waren [1]. Daten mit gesicherten Außenkriterien konnten von 242 Fällen gesammelt werden. Die Untersuchten gehören zu folgenden 6 Gruppen:

1. Normalfälle (32 Fälle),
2. Patienten mit Pankreaskarzinom (21 Fälle),
3. Patienten mit calzifizierender Pankreatitis (41 Fälle),
4. Patienten mit chronischer Pankreatitis (64 Fälle),
5. Patienten mit akutem Magen- oder Duodenalulcus (36 Fälle),
6. Patienten mit chronischem Ulcus (48 Fälle).

### Diskriminanzanalyse

Bei der Gegenüberstellung der Mittelwerte und der Standardabweichungen der einzelnen Enzymfraktionen bei Gesunden



und bei Patienten mit Pankreaskarzinom erkennt man jeweils in der 1. Fraktion nach einem Pankreasreiz die stärksten Differenzen der Mittelwerte. Die Mittelwerte der übrigen Gruppen liegen zwischen diesen beiden Extremen.

Die lineare Diskriminanzanalyse zwischen Normalfällen und Patienten mit einer der 3 Pankreaserkrankungen zeigt eine weitgehende Trennung der beiden Gruppen und eine nur geringfügige Überlappung der beiden Verteilungen aufgrund des Mahalanobis Abstands [2].

Bei der Erarbeitung der ersten Trennformeln war es von großer Wichtigkeit, zu einer Reduktion der 77 Variablen zu kommen. Gemäß den Erfahrungen der Klinik aus der monovariablen Betrachtungsweise der Variablen, wurden jeweils die Daten aus der 1. Fraktion nach einem Pankreasreiz ausgewählt, wodurch die 77 Variablen auf 26 reduziert werden konnten [3]. Innerhalb dieser 26 Variablen wurden durch eine aufbauende lineare Diskriminanzanalyse die trennfähigsten 13 Variablen ausgewählt, weil darüberhinaus durch Hinzunahme weiterer Variablen der Mahalanobis Abstand nicht mehr zunahm. Es wurden Werte von Trypsin und Lipase sowie Saftvolumen und Bicarbonat und einmal die Bilirubinkonzentration ausgewählt.

Die nichtlineare Diskriminanzanalyse erreichte bei der Kreuzvalidierung eine Trefferrate von 46 % für 6 Gruppen (Tab. 1) [4]. In der Zuordnungsmatrix fiel auf, daß bestimmte Felder um die Diagonale der richtigen Zuordnung vermehrt besetzt waren. Durch Zusammenfassung verwandter Krankheiten zu den 3 Gruppen: 1. Normalfälle, 2. Patienten mit Pankreaserkrankungen und 3. Patienten mit Ulcuserkrankungen erhöht sich die Trefferrate auf 77 %. Durch eine weitere Zusammenfassung auf 2 Gruppen (Fälle mit Pankreaserkrankungen und Fälle ohne Pankreaserkrankungen) erreicht die Trefferrate schließlich 96 %. Der 3-Gruppenfall zeigt, daß jetzt eine Mittelgruppe (Patienten mit Ulcus) zwischen Patienten mit einer Pankreaserkrankung und den Normalfällen abgetrennt werden konnte. Dieses Ergebnis änderte sich kaum durch Hinzunahme weiterer Krankheiten, die in der Umgebung des Pankreas liegen und auf die Saftsekretion mindernd einwirken, d.h. die Umgebungserkrankungen verhalten sich im wesentlichen wie die Ulcuserkrankungen.

Die Diskriminanzanalyse wird seit Oktober 1973 an der Klinik routinemäßig als Diagnostikunterstützung eingesetzt. Die Daten der täglich durchgeführten Pankreasfunktionstests werden seither durch die Diskriminanzformel neu eingeordnet und das Ergebnis zur Diagnostik herangezogen. Mittlerweile wurde die Diskriminanzformel durch Hinzunahme

der Variablen Alter, Gewicht und Geschlecht und durch Vergrößerung der Gruppenumfänge verbessert. Die Klinik führt jetzt die Neueinordnung selbständig auf ihrem Laborcomputer durch. Das Institut liefert lediglich jährlich die neue Diskriminanzformel. Die Trefferraten der Neueinordnung betragen im 6-Gruppenfall 65 %, im 3-Gruppenfall 86 % und im 2-Gruppenfall 98 % [4]. Gegenüber den Beurteilungsmöglichkeiten durch den Arzt haben sich die Ergebnisse erheblich verbessert, dessen Trefferrate im 2-Gruppenfall bei 82 %, im 3-Gruppenfall bei 38 % liegt, und der über den 6-Gruppenfall keine Aussage machen kann.

### Clusteranalyse

Um die Leistung eines statistischen Verfahrens abschätzen zu können, das ohne Außenkriterien arbeitet, wurden die Daten (13 Variablen) einer Clusteranalyse (Programmpaket CLUSTAN – Version 1c – WISHART) unterzogen [6, 7]. Die eigentliche Klassifizierung wurde mit der Routine HIERARCHY nach der WARD-Methode gerechnet, in der die Summe der Fehlerquadrate minimiert wird. Als Ergebnis bekommt man erstens die Daten für ein Dendrogramm, das die einzelnen Fusionsschritte graphisch darstellt, und zweitens eine Klassifikationsmatrix, die als Eingabe für das Cluster- oder für das Scatterdiagramm dient. Bei dieser Darstellungsweise werden die Fälle in eine Ebene projiziert, wobei die beiden Achsen Variablen, Hauptfaktoren oder eine Kombination von beiden sein können. Die Längen dieser Achsen sind proportional der jeweiligen Standardabweichung in der entsprechenden Richtung. Beim Scatterdiagramm wird die Position eines Falles durch ein Kreuz (+) und wahlweise zusätzlich durch eine Identifikation markiert. Im vorliegenden Fall wurden die tatsächliche Diagnose und das Ergebnis der Diskriminanzanalyse in die Identifikation mit aufgenommen. Im Clusterdiagramm (Abb. 1) werden die Cluster einer beliebig wählbaren Ebene durch Kreise dargestellt; Mittelpunkt eines Clusterkreises ist der Schwerpunkt des Clusters und sein Radius ist proportional der Wurzel aus der Summe ihrer Varianzen in den Achsenrichtungen. Jeder Clusterkreis wird durch einen Code aus der normalisierten Klassifikationsmatrix im Mittelpunkt des Clusterkreises markiert. Als Alternative können alle Fälle in das Clusterdiagramm projiziert werden, wobei die Positionen der Fälle eines Clusters mit der Codenummer des Clusters gekennzeichnet werden (Abb. 2).

Es wurden die Klassifikationsmatrizen für die Daten des

Tab. 1. Einordnung von 242 Patienten durch eine nichtlineare Diskriminanzanalyse

Einordnung durch Diskriminanzanalyse \ Tatsächliche Diagnose	Diagnosen						Summe
	Pankreaskarzinom	Calzif. Pankreatitis	Chronische rezidiv. Pankreatitis	Akutes Ulkus	Chronisches Ulkus	Normale Pankreasfunktion	
Pankreaskarzinom	5	6	4	1			16
Calzifizierende Pankreatitis	8	22	19	1			50
Chronisch rezid. Pankreatitis	6	9	30	4	3	1	53
Akutes Ulkus ventr. et Duodeni		3	2	8	7	3	23
Chron. Ulkus ventr. et Duodeni	2	1	7	15	31	12	68
Normaler Pankreasfunktionstest			2	7	7	16	32
Summe	21	41	64	36	48	32	242



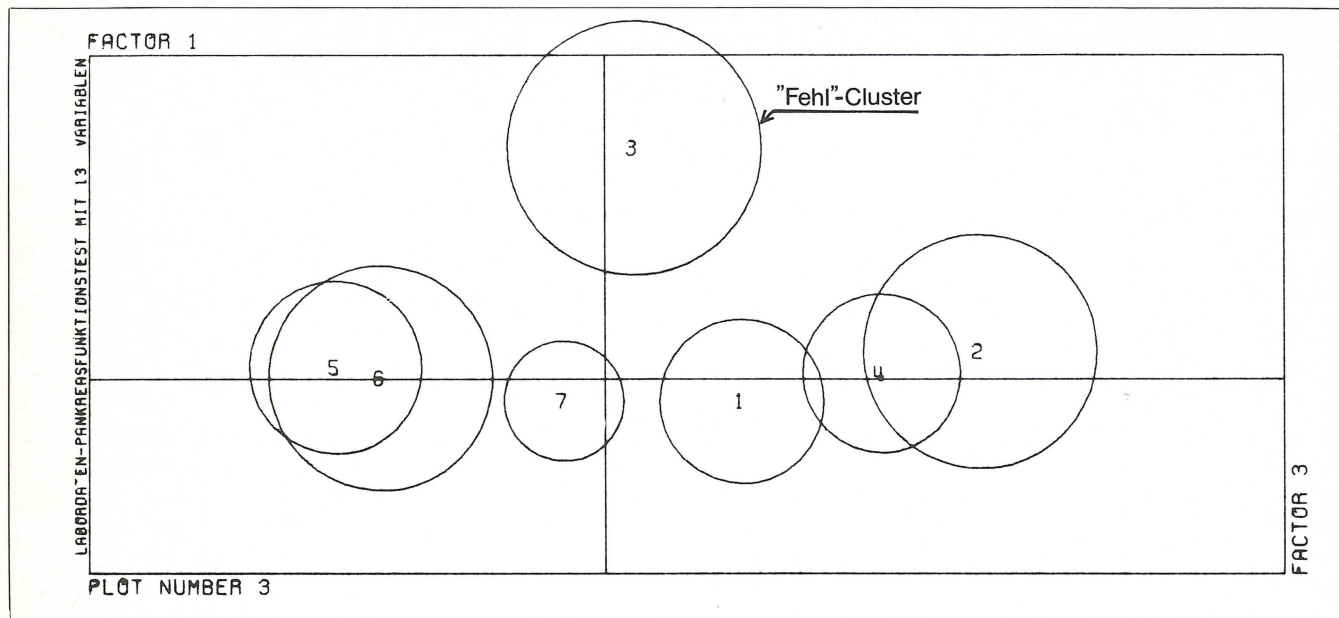
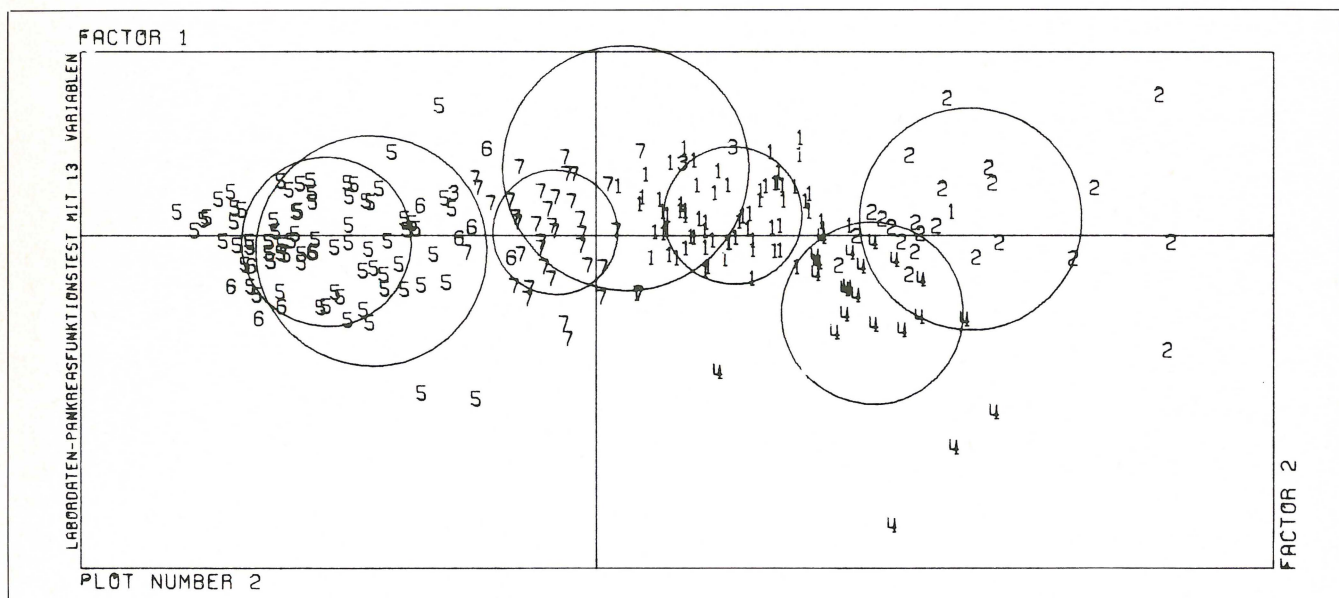


Abb. 1. Clusterdiagramm. Die Kreise stellen die Lage der Cluster im Koordinatensystem des 1. und 3. Hauptfaktors dar. Sie werden durch die Code der Klassifikationsmatrix symbolisiert.

Pankreasfunktionstestes von den letzten 10 bis herunter zu den letzten 2 Clustern erzeugt. Um die Frage: »Können die 6 Diagnosegruppen in Form von Clustern wiedergefunden werden?« zu beantworten, wurde zunächst eine Gegenüberstellung der Diagnosegruppen mit den letzten 6 Clustern in einer  $6 \times 6$  Matrix vollzogen (Tab. 2). Da nur »sinnvolle« Cluster betrachtet werden sollten, wurde ein Cluster, der lediglich aus 3 Vertretern bestand, die außerdem sehr abseits vom übrigen Datenmaterial lagen, als »Fehl-Cluster« bezeichnet. Dadurch wurden die letzten 7 Cluster abzüglich dieses »Fehl-Clusters« zu den letzten 6 Clustern erklärt.

Abb. 2. Clusterdiagramm mit Projektion der Fälle. In das Clusterdiagramm (Hauptfaktor 1 und 2) sind alle Fälle eingetragen und ihre Clusterzugehörigkeit mit der Codenummer des Clusters gekennzeichnet.



Die Zuordnung der Cluster zu einer bestimmten Diagnose konnte nun nicht nach der häufigsten Diagnose, die ein Cluster enthielt, vorgenommen werden, da gelegentlich eine Diagnose in 2 Clustern am häufigsten auftrat. Die Cluster wurden daher so zugeordnet, daß die Trefferrate der gesamten Matrix am höchsten war (Tab. 2). Bei dieser Vorgehensweise ist allerdings der Cluster für die Karzinomfälle sehr schwach besetzt. Die Trefferrate der Clusteranalyse beträgt demnach 36 %, wobei wieder der 3-Gruppen- (65 %) und der 2-Gruppenfall (88 %) nahelag.

Im Clusterdiagramm konnte nun die Bezeichnung der Cluster eingetragen werden. Damit zeigt sich, daß die Cluster von den Patienten mit calcifizierender Pankreatitis bis zu den Normalfällen in Richtung des 1. Hauptfaktors angeordnet sind und zwar in der Reihenfolge: calcifizierende Pankreatitis, Pankreaskarzinom, chronische Pankreatitis, akutes Ulcus, chronisches Ulcus und Normalfälle. In Richtung des 2. bzw. des 3. Hauptfaktors ist kein signifikanter Trennunterschied zu erkennen. Die Wichtigkeit der Hauptfaktoren geht aus ihren Parametern hervor (Tab. 3).



Tab. 2. Einordnung von 242 Patienten durch eine Clusteranalyse

Tatsächliche Diagnose Interpretation und Cluster-Nr.	Diagnosen						Summe
	Pankreas- karzinom	Calzif. Pankreatitis	Chron. rezid. Pankreatitis	Akutes Ulkus	Chronisches Ulkus	Normale Pankreas- funktion	
Pankreaskarzinom (Cluster 6)	1	5	4				10
Calzifizierende Pankreatitis (Cluster 5)	15	28	33				76
Chronisch rezid. Pankreatitis (Cluster 7)	3	5	18	8	4		38
Akutes Ulkus ventr. et Duodeni (Cluster 1)	2	3	4	19	26	16	70
Chron. Ulkus ventr. et Duodeni (Cluster 4)			3	2	12	5	22
Normaler Pankreasfunktionstest (Cluster 2)				7	6	10	23
Fehler (Cluster 3)			2			1	3
Summe	21	41	64	36	48	32	242

### Faktorenanalyse

Um die Faktoren deuten zu können, wurde zusätzlich eine Faktorenanalyse mit 13 Variablen aus den beiden 1. Fraktionen gerechnet [8]. Die Faktorenanalyse mit orthogonaler Rotation lieferte 4 Faktoren, die 80 % der Varianz erklären. In der Matrix der Faktoren-Ladungen aus dem Programm BMDP4M werden die Zeilen so umgeordnet, daß in jedem Faktor die hohen Ladungen am Anfang stehen. Gleichzeitig sind Ladungen unter 0,25 durch Nullen ersetzt. Auf diese Weise wird die Interpretation der Faktoren stark erleichtert, wenn die Variablen zusätzlich durch Kurzbezeichnungen gekennzeichnet sind.

Nach dieser Ladungsmatrix kann der 1. Faktor als die ekbolische Funktion des Pankreas interpretiert werden, die sich bekanntermaßen am Enzymauswurf des Pankreas zu erkennen gibt. Der 2. Faktor repräsentiert die Bikarbonatmenge und das Trypsin, und der 3. Faktor drückt sich in erster Linie im Saftvolumen aus.

### Ergebnisse

Unter Rückbezug auf das Clusterdiagramm kann man aus der Clusteranalyse erkennen, daß die Diagnosegruppen nach der ekbolischen Funktion angeordnet sind (diskriminierende Achse). Anders ausgedrückt: Die Erkrankungen stören das Pankreas vorwiegend in seiner ekbolischen Funktion. Erkrankungen der Umgebung (Ulcus des Magens oder des Zwölffingerdarms) schränken diese Pankreasfunktion in viel geringerem Maße ein, wobei das akute Ulcus sich stärker auswirkt als das chronische. Die statistische Analyse bestätigt hier den Eindruck, den die Klinik im Laufe der Zeit gewonnen hat. Im Scatterdiagramm erkennt man, daß sich die Diagnosegruppen viel stärker überlappen, als das im Clusterdiagramm sichtbar wird. Ausgehend vom Stand der letzten 6

Cluster kann nun das Clustern in zwei Richtungen verfolgt werden:

1. Wie vereinen sich die Gruppen weiter?
2. Wie entstanden diese Gruppen?

### Betrachtung des Fusionsverlaufs der Clusteranalyse

Die 6 Cluster vereinigen sich weiter so, daß zunächst die Gruppe des chronischen Ulcus und die Normalfälle in eine Gruppe mit ungestörter bzw. leicht gestörter Pankreasfunktion fallen. Dann schließen sich die Gruppen des akuten Ulcus und die der chronischen Pankreatitis zu einer Gruppe von mittelschwerer Pankreasfunktionsstörung zusammen (Abb.3). Schließlich vereinigen sich die Gruppen der calzifizierenden Pankreatitis und die des Pankreaskarzinoms zur Gruppe der schweren Pankreasfunktionsstörung. Dieser 3-Gruppenfall entspricht dem 3-Gruppenfall der Diskriminanzanalyse nur teilweise, da die Grenze der Gruppen aus der Clusteranalyse jeweils um 1 Diagnose gegenüber der Diskriminanzanalyse verschoben ist. Im 2-Gruppenfall ist die Grenze ebenfalls gegenüber der Diskriminanzanalyse verschoben. Dagegen stellt sich der 2-Gruppenfall entsprechend der Diskriminanzanalyse im Clusterdiagramm dar, da die Cluster mit Pankreaserkrankungen im negativen Bereich und die Cluster ohne Pankreaserkrankungen im positiven Bereich des 1. Hauptfaktors liegen.

Verfolgt man die Entstehung der Cluster in den Ebenen 10 bis 7, so zeigt sich, daß die Cluster der Normalfälle, des akuten Ulcus und der calzifizierenden Pankreatitis aus der Fusionierung von je 2 Clustern entstehen. Auf dem entsprechenden Clusterdiagramm kann man erkennen, daß sich die Lage der verschmelzenden Gruppen vorwiegend in Richtung der 2. bzw. 3. Hauptachse unterscheidet.

Betrachtet man nun die Übereinstimmung zwischen Clusteranalyse und Diskriminanzanalyse in Bezug auf die tatsächlichen Verhältnisse, so werden von beiden Verfahren gleichzeitig nur 45 von den 242 Patienten richtig (19 %) und 89 Patienten (37 %) falsch eingeordnet. Markiert man im Clusterdiagramm die Fälle mit den Einordnungsergebnissen der beiden Verfahren in Bezug auf die tatsächliche Diagnose, so sieht man eine Reihe von benachbarten Fällen, bei denen beide Verfahren gleiche Ergebnisse liefern.

Tab. 3. Parameter der Hauptfaktoren

Hauptfaktor	1	2	3
Parameter			
Eigenwert	7.41	1.40	1.10
Prozent der Gesamtvarianz	57.00	10.78	8.43
Kumulative Varianz	57.00	67.78	76.21



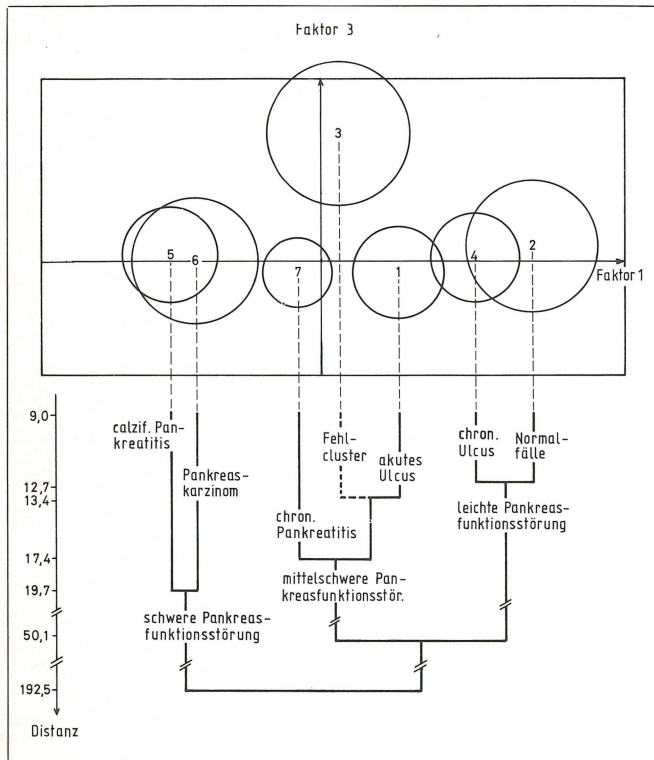


Abb. 3. Clusterdiagramm mit Dendrogramm. Das Clusterdiagramm ist durch das Dendrogramm der letzten 7 Cluster ergänzt.

Abschließend betrachtet, liefert die Clusteranalyse die zugrundeliegenden Diagnosegruppen, wenn man auf einer Ebene verharret, auf der sich die Gruppen im Clusterdiagramm als nicht überschneidende Gruppen darstellen. Das ist der 3-Gruppenfall, der sich auch in der Diskriminanzanalyse als arbeitsfähige Gruppeneinteilung erweist. Eine richtige Einordnung der Einzelfälle in diese Gruppen wird jedoch nicht in ausreichendem Maße erreicht, wenn auch die Möglichkeiten übertroffen werden, die der Arzt bei der subjektiven Beurteilung der Untersuchungsdaten hat. Im vorliegenden Fall

brachte die Clusteranalyse zusammen mit dem Clusterdiagramm eine Abstufung der Diagnosen in Richtung des 1. Hauptfaktors, der mit Hilfe der Faktorenanalyse als Beeinträchtigung der ekbolischen Pankreasfunktion gedeutet werden konnte, ein Ergebnis, das mit der Diskriminanzanalyse nicht gefunden wurde.

## Literatur

- [1] THURMAYR, R.: Computergerechte Dokumentation des Pankreasfunktionstestes und deren Ergebnisse. Bericht MD 136 (1975) Ges. f. Strahlen- und Umweltforschung mbH München, Ingolstädter Landstr. 1, 8042 Neuherberg
- [2] OTTE, M., THURMAYR, R., THURMAYR, G. R.: Computerhilfe bei der Pankreasfunktionsdiagnostik. Diagnostik 5, 656 – 660 (1972)
- [3] THURMAYR, R., THURMAYR, G. R., OTTE, M.: Probleme beim Routineinsatz der Diskriminanzanalyse zur Beurteilung des Pankreasfunktionstestes. EDV in Medizin und Biologie, 6, 1/2, 49 – 52 (1975)
- [4] VICTOR, N., HÖRMANN, A., EDER, L.: STATSYS Beschreibung und Benutzeranleitung. Bericht MD 24 (1973) Ges. f. Strahlen- und Umweltforschung mbH München, Ingolstädter Landstr. 1, 8042 Neuherberg
- [5] THURMAYR, R., BLOMER, R. J., FORELL, M. M., JAFFÉ, A., OTTE, M., RASCHEWA, C., THURMAYR, G. R.: Computer aided diagnosis of pancreatic function tests in the routine situation. In: Decision making and medical care. Can information science help? Proceedings of the IFIP Working Conference on Decision Making and Medical Care, 24. – 29. 5. 1976, Dijon, France (Eds.: F. T. de Dombal and F. Gremy) Amsterdam: North-Holland Publishing Company, 175 – 183 (1976)
- [6] WISHART, D.: Clustan User Manual. Inter-University / Research Councils Series Report No. 47, Edinburgh (1978)
- [7] WARD, J. H.: J. Amer. Stat. Ass. 58, 236 (1963)
- [8] DIXON, W. J.: Biomedical Comput Programms, University of California Press, Berkeley, Los Angeles, London (1975)

Anschrift der Verfasser: Dr. J. B. Schäffer, Institut für Med. Datenverarbeitung der GSF, München-Neuherberg, Arabellastr. 4/III, 8000 München 81; Prof. Dr. med. R. Thurmayer und Dr. med. G. R. Thurmayer, Institut für Med. Statistik und Epidemiologie der Fakultät für Medizin der TU München, Sternwartstr. 2/II, 8000 München 80; Prof. Dr. med. M. Otte, Gastroenterologische Abteilung der Med. Klinik Innenstadt, Ziemssenstr. 1, 8000 München 2

EDV in Medizin und Biologie 10 (2), S. 58 – 61, ISSN 0300-8282  
© Eugen Ulmer GmbH & Co., Stuttgart; Gustav Fischer Verlag KG, Stuttgart

## Sample size determination by simulation

K.-H. Jöckel

### Zusammenfassung

Das hier vorgeschlagene Simulationsverfahren ermöglicht die Bestimmung des notwendigen Stichprobenumfangs zur Erreichung einer vorgegebenen Güte für spezielle Alternativen. Es erweist sich, daß das Verfahren asymptotisch gute Eigenschaften besitzt. An Hand von Beispielen wird eine gute Übereinstimmung mit exakten Ergebnissen und eine Überlegenheit gegenüber asymptotischen Approximationen gezeigt.

### Summary

A simulation procedure is proposed which allows the determination of sample sizes for achieving a preassigned power at

specific alternatives. The procedure is shown to have good asymptotic properties. Examples further demonstrate a good accordance with exact results and a superiority to asymptotic approximations.

### 1. Introduction

Let  $(\Phi_n)_{n \in \mathbb{N}}$  be a sequence of level  $\alpha$ -tests at sample size  $n$  for testing  $H_0$  against  $H_1$ . In nearly all practical investigations the following question arises: How large has to be the sample size  $n$  such that the chosen test  $\Phi_n$  reaches a certain power say  $\beta \in (0,1)$  against a fixed alternative  $(P_n)_{n \in \mathbb{N}}$ . Thus we are interested in

$$(1.1) \quad N = \inf \{n \in \mathbb{N} : E(\Phi_n | P_n) \geq \beta\}, \text{ where } \inf \Phi = \infty.$$



For many test problems this question can be solved analytically. But especially in nonparametrics there is usually no closed form for  $E(\Phi_n|P_n)$ . Normally this problem is attacked by asymptotic approximations of the power functions. Here we want to propose a different solution: estimate the  $N$  above by simulation. The procedure then is the following. For every  $n$  we have a natural estimator for  $E(\Phi_n|P_n)$ , say  $S_m^n$  based on  $m$  Monte-Carlo runs. We set,

$$(1.2) \quad N_m = \inf \{n \in \mathbb{N} : S_m^n \geq \beta\}$$

and  $N_m$  serves as an estimator for  $N$ .

In the following section 2 we shall establish a suitable probabilistic model for studying the asymptotic properties of the estimator in section 3. The theorems of 3 can be understood in terms of this introduction. Thus the reader who is primarily interested in applications may omit 2 and directly go on to 3. In section 4 we give some examples which demonstrate the use and power of the procedure.

## 2. The Model

For any  $n \in \mathbb{N}$  let  $\Phi_n$  be a test on a sample space  $(X_n, B_n)$  and  $P_n$  a probability measure on  $(X_n, B_n)$ . Starting with  $m$  independent realizations  $x_1, x_2, \dots, x_m$  of a random variable distributed according to  $P_n$  we define for each  $n$  an estimator for  $E(\Phi_n|P_n)$  by

$$(2.1) \quad \tilde{S}_m^n(x_1, \dots, x_m) = m^{-1} \sum_{i=1}^m \Phi_n(x_i)$$

on  $(X_n^m, B_n^m, P_n^m)$ , where the latter probability space means the  $m$ -fold product of  $(X_n, B_n, P_n)$ . This estimator is unbiased and according to the strong law of large numbers consistent, i.e.

$$\lim_{m \rightarrow \infty} \tilde{S}_m^n = E(\Phi_n|P_n) \quad \text{a.e.}$$

Let  $\beta \in (0,1)$  be fixed and  $N = \inf \{n \in \mathbb{N} : E(\Phi_n|P_n) \geq \beta\}$  the parameter of interest, then the probability space

$$(\Omega_m, \mathcal{A}_m, \mathbb{P}_m) = \left( \prod_{n=1}^{\infty} X_n^m, \prod_{n=1}^{\infty} B_n^m, \prod_{n=1}^{\infty} P_n^m \right)$$

is the adequate model. If  $m$  is fixed the subscript  $m$  will be suppressed. Denoting by  $\pi_n$  the projection from  $(\Omega, \mathcal{A}, \mathbb{P})$  onto the  $n$ -th factor space  $(X_n^m, B_n^m, P_n^m)$  we define

$$(2.2) \quad S_m^n = \tilde{S}_m^n \circ \pi_n$$

which again is an unbiased and consistent estimator for  $E(\Phi_n|P_n)$  but now defined on the overall probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ .

Let

$$(2.3) \quad N_m = \inf \{n \in \mathbb{N} : S_m^n \geq \beta\}$$

which is a measurable function  $(\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\overline{\mathbb{N}}, 2^{\overline{\mathbb{N}}})$  ( $2^{\overline{\mathbb{N}}}$  denoting the power set of  $\mathbb{N} = \overline{\mathbb{N}} \setminus \{\infty\}$ )

## 3. Asymptotic Properties of the Estimator

Under mild additional assumptions on the test problem we shall show that the proposed procedure has nice asymptotic properties.

**THEOREM 1:** Let  $(\Phi_n)_{n \in \mathbb{N}}$  be a sequence of tests such that there exists a positive integer  $n_0$  and a real  $\delta$  with

$$(3.1) \quad E(\Phi_n|P_n) \geq \delta > \beta \text{ for all } n \geq n_0.$$

Then  $N_m$  is finite with probability 1 and has finite expectation.

**Proof:** Using the unbiasedness of  $S_m^n$  and (3.1) we get for all  $n \geq n_0$

$$\begin{aligned} \delta \leq E(S_m^n) &= \int_{[0, \beta)} x \, d\mathbb{P}^{S_m^n} + \int_{[\beta, 1]} x \, d\mathbb{P}^{S_m^n} \\ &\leq \beta \mathbb{P}(S_m^n < \beta) + 1 - \mathbb{P}(S_m^n < \beta) \\ &= 1 - (1-\beta) \mathbb{P}(S_m^n < \beta). \end{aligned}$$

Consequently

$$(3.2) \quad \mathbb{P}(S_m^n < \beta) \leq (1-\delta)(1-\beta)^{-1} = q < 1 \text{ for all } n \geq n_0.$$

According to (2.3) we have

$$(3.3) \quad \mathbb{P}(N_m = k) = \left[ \prod_{i=1}^{k-1} \mathbb{P}(S_m^i < \beta) \right] \cdot \mathbb{P}(S_m^k \geq \beta)$$

This together with (3.2) yields

$$\mathbb{P}(N_m = \infty) = 0 \text{ and}$$

$$EN_m \leq \sum_{k=1}^{n_0} k \mathbb{P}(N_m = k) + \sum_{k=n_0+1}^{\infty} k q^{k-n_0} < \infty.$$

**Remarks:** 1) The assumption (3.1) in Theorem 1 is e.g. fulfilled if  $\lim_{n \rightarrow \infty} E(\Phi_n|P_n) = 1$ , that's to say, if  $(\Phi_n)_{n \in \mathbb{N}}$  is a con-

sistent sequence of tests.

2) Easy calculations show that in fact all higher moments of  $N_m$  exist in the situation of Theorem 1.

**THEOREM 2:** Let  $E(\Phi_n|P_n)$  be a weakly increasing sequence in  $n$ ,  $\lim_{n \rightarrow \infty} E(\Phi_n|P_n) > \beta$  and  $E(\Phi_n|P_n) \neq \beta$  for all  $n$ . Then  $N_m$

is asymptotically unbiased, i.e.  $\lim_{m \rightarrow \infty} EN_m = N$

( $N$  defined by (1.1)).

**Proof:** There exists a  $\delta > \beta$  such that

$$E(\Phi_n|P_n) \geq \delta > \beta \text{ for all } n \geq N.$$

Because of  $E(S_m^n) - \beta = E(\Phi_n|P_n) - \beta \geq \delta - \beta > 0$  for all  $n \geq N$  we get using Tschebychev's inequality

$$\begin{aligned} \mathbb{P}(S_m^n < \beta) &\leq \mathbb{P}(|S_m^n - E(S_m^n)| > E(S_m^n) - \beta) \\ &\leq (\sigma - \beta)^{-2} \text{Var } S_m^n = m^{-1} (\delta - \beta)^{-2} \text{Var } \Phi_n \\ &\leq m^{-1} (\delta - \beta)^{-2} (1 - \delta) = q_m. \end{aligned}$$

Consequently

$$(3.4) \quad EN_m \leq \sum_{k=1}^N k \mathbb{P}(N_m = k) + \sum_{k=N+1}^{\infty} k \cdot q_m^{k-N}.$$

Since  $E(\Phi_n|P_n) \neq \beta$  for all  $n$  and  $S_m^n$  is consistent the first term of the right hand side of (3.4) tends to  $N$  as  $m$  tends to infinity. As the second term converges to 0, we conclude  $\lim_{m \rightarrow \infty} EN_m \leq N$ .

Obviously  $EN_m \geq \sum_{k=1}^N k \mathbb{P}(N_m = k) \rightarrow N$ ,  $m \rightarrow \infty$ , which yields the assertion.



Remark: Under the conditions of Theorem 2 easy calculations yield  $\text{Var } N_m = O(m^{-1})$

In the sequel we shall also consider the case  $E(\Phi_n | P_n) = \beta$  for some  $n$ , always assuming  $\Phi_n \neq \text{const.}$   $P_n - \text{a.e.}$ , that's to say  $\text{Var } \Phi_n \neq 0$ .

THEOREM 3: Let  $E(\Phi_n | P_n)$  be a weakly increasing sequence with  $\lim_{n \rightarrow \infty} E(\Phi_n | P_n) > \beta$ .

a) If  $\{n \in \mathbb{N} : E(\Phi_n | P_n) = \beta\} = \emptyset$ , we have

$$\text{IP}(N_m = k) \rightarrow \begin{cases} 1 & k = \inf \{n \in \mathbb{N} : E(\Phi_n | P_n) \geq \beta\} = N \\ 0 & \text{otherwise} \end{cases} \quad m \rightarrow \infty$$

b) If  $\{n \in \mathbb{N} : E(\Phi_n | P_n) = \beta\} = \{N, N+1, \dots, N+q\}$ ,  $q \in \mathbb{N} \setminus \{0\}$ , then

$$\text{IP}(N_m = k) \rightarrow \begin{cases} 2^{-(r+1)} & k = N + r, r \in \{0, \dots, q\} \\ 2^{-(q+1)} & k = N + q + 1 \\ 0 & \text{otherwise} \end{cases} \quad m \rightarrow \infty$$

Proof: a) Applying Tschebychev's inequality as in the proof of Theorem 2 it is easy to see, that  $\text{IP}(N_m = k)$  tends to 1 as  $m \rightarrow \infty$ , if  $k = N$  and to 0 otherwise.

b) For  $k \notin \{N, N+1, \dots, N+q+1\}$  everything is clear. Let  $k \in \{N, \dots, N+q+1\}$ , then

$$\text{IP}(S_m^k < \beta) = \text{IP}\left((m \text{Var } \Phi_k)^{-1/2} \sum_{i=1}^m (\Phi_k(x_i) - \beta) < 0\right)$$

$$\text{IP}(S_m^k \geq \beta) = \text{IP}\left((m \text{Var } \Phi_k)^{-1/2} \sum_{i=1}^m (\Phi_k(x_i) - \beta) \geq 0\right)$$

both expressions converging to 1/2 according to the central limit theorem, which finishes the proof.

Remark: A consequence of Theorem 3 b) is that

$EN_m \rightarrow N + 2^{-(q+1)}(2^{q+1} - 1)$  which shows that the asymptotic bias is not too large.

#### 4. Examples and some Comparisons

In the following section we want to show the effectiveness of the described simulation procedure by comparison with exact values and to demonstrate its superiority to asymptotic approximative solutions. That's why we apply it to some standard unsophisticated test situations.

EXAMPLE 1: Given  $n$  independent identical normally distributed random variables with variance 1 and mean  $\mu$  unknown we want to test  $H_0 : \mu \leq 0$  against  $H_1 : \mu > 0$  at level 0.01.

The test is intended to detect a difference in mean of 0.9 with an error-II-probability of at most 10 %. How many observations  $N$  have to be taken? Of course the answer is trivial. The appropriate test is the one-sided one-sample-Gauß-test, see e.g. (Witting, 1970) p. 186. The power function of this test is well known. Easy calculations yield  $N$  to be 9, achieving a power of 92 %. Table I shows that the estimation procedure has worked very well.

In the sequel we want to consider test situations where closed expressions for the power functions are no longer available. As an example we take the Wilcoxon two-sample-rank-test which will be applied under two different distribution assumptions. But before doing so we have to say something according to the two-sample (or  $k$ -sample) situation. We want to determine the sample size to achieve a certain power. But what does sample size in this context actually mean? Of course the size of the pooled sample. This, however, requires a predetermined fixed sampling plan, that's to say a relationship between the single sample sizes and the size of the pooled sample. Generally this will present not too big a problem, because the experimenter will often have a preknowledge of the variability of the sample (when being concerned with shift models).

If the samples have different costs of observation another factor comes into the analysis. In this case repeated runs of the simulation procedure with different sampling plans allow us to minimize (at least partially) the costs of the study under the condition of a preassigned power.

EXAMPLE 2: As has already been said no closed form for the power function of the two-sample Wilcoxon test is available. But fortunately in the case of a normal shift model there do exist tables (Milton, 1970) which give some selected exact power results. That is one reason for having chosen this example.

We start with two samples of independent random variables  $X_1, \dots, X_{n_1}, Y_1, \dots, Y_{n_2}$  where the  $X_i$ 's respectively the  $Y_j$ 's are identically distributed  $i = 1, \dots, n_1, j = 1, \dots, n_2$ . A location model might be adequate and we want to test  $H_0 : \Delta \leq 0$  against  $H_1 : \Delta > 0$  at level  $\alpha = 0.05$ , where  $\Delta$  denotes the location parameter. We shall use the Wilcoxon rank-sum-test. It is assumed that we want to detect a location difference of  $\Delta = 1$  with a probability of at least 50 % under the condition that the underlying distribution is normal with fixed variance  $\sigma^2 > 0$ . The test will then detect even any larger difference, see e.g. (Witting, 1970) pp. 118, 119, Theorem 3.13. As the power of the test only depends on the ratio  $\Delta/\sigma$  we put  $\sigma^2 = 1$ . Having chosen a shift model we take both sample sizes  $n_1, n_2$  to be equal. Since the level is 0.05 ran-

Tab. 1. Sample Size Determination for the One-Sample Gauß-Test for  $IN(\mu, 1), \mu = 0.9, \alpha = 0.10$

Sample Size $n$	3	4	5	6	7	8	9
Simulated Power $S_n^{10000}$	0.6143	0.6998	0.7711	0.8217	0.8632	0.8957	0.9180
$N_{10000} = 9$							

Tab. 2. Sample Size Determination for the Wilcoxon Test in the Normal Case,  $\Delta = 1, \alpha = 0.05$

Sample Size 1 <sup>st</sup> Sample $n_1$	3	3	4	4	5	5	6	6	7
Sample Size 2 <sup>nd</sup> Sample $n_2$	3	4	4	5	5	6	6	7	7
Simulated Power $S_{n_1+n_2}^{10000}$	0.2612	0.2946	0.3310	0.3728	0.4055	0.4251	0.4549	0.4913	0.5176
$N_{10000} = 14$									



Tab. 3. Sample Size Determination for the Wilcoxon Test under Exponentiality  $\Delta = 1$ ,  $\alpha = 0.05$ 

Sample Size 1 <sup>st</sup> Sample $n_1$	3	3	4	4	5	5	6
Sample Size 2 <sup>nd</sup> Sample $n_2$	3	4	4	5	5	6	6
Simulated Power $S_{n_1+n_2}^{10000}$	0.3874	0.3880	0.4576	0.4969	0.5463	0.5850	0.6263
$N_{10000} = 12$							

domization is necessary. But of course we might have taken only the levels actually attained according to the generality of the model. The procedure gives  $N = N_1 + N_2 = 7 + 7$  (see Table II) where  $N_1, N_2$  denotes the size of the first, second sample. The estimated power is 0.52. Again these values are in good agreement with (Milton, 1970) Table 3 on page 34, which shows the exact power to be 0.525.

Now we come to the discussion of asymptotic approximations. Using e.g. the method described in (Lehmann, 1975) pp. 72–74 we compute the necessary sample size as  $6 + 6 = 12$ , the approximative power being 0.493. Table 3 in (Milton 1970) shows that the exact power is only 0.466. In this case the result of the approximative solution is not too bad. The next example, however, will show that our procedure is essentially better.

**EXAMPLE 3:** In the terminology of Example 2 the  $X_i$ 's are assumed to be exponentially distributed with parameter  $\lambda > 0$ , that's to say the distribution function is  $(1 - \exp(-\lambda x))$  for  $x \geq 0$ , zero otherwise. Again the  $Y_j$ 's are shifted by an amount  $\Delta$  which is unknown. Using the Wilcoxon test at level  $\alpha = 0.05$  we want to detect a  $\Delta = 1$  with a probability of at least 60 %, where the scale parameter  $\lambda$  has been set to 1. Table III gives a sample size of 12 for the pooled sample, consequently 6 in each sample. Now the use of the approximation formula in (Lehmann, 1975) p. 72 yields a sample size of only  $3 + 3 = 6$  and an approximative power of 0.62. The estimated power, however, is then only 0.39. Other simulations by the author have shown that at a sample size of 6 60%-power is reached only at a location difference of more than 1.5. This shows that the use of approximations may produce considerable mistakes.

## 5. Conclusions and Prospect

The procedure proposed in this paper should be used whenever there is no closed expression for the power function at hand. As the examples show it dominates approximative solutions. And what is more important, it is applicable also in those cases where no approximative solutions are known, even if

no exact determination of the critical regions is possible, if only the power function is really monotone increasing in sample size.

Another advantage is the following: The procedure allows to make some inference about the accuracy of the estimation, for details see 6, whereas for the approximative solution there are generally no error estimates available.

Besides application in experimental design the procedure might be useful for investigating empirical efficiencies. In this case but also in the former one there might arise the problem of too many Monte-Carlo runs. Thus it seems worthwhile to study the estimator when using dependent trials. This and a more elaborate distribution theory will be done in a forthcoming paper.

## 6. Details of Simulation

The simulation study was executed on the IBM-Computer of the HRZ-Dortmund. All programs were written in Fortran. The random numbers were generated by transforming uniformly distributed random numbers of RANDU (IBM, 1970) by the specific inverted distribution function. The use of 10000 trials for the simulation yields 90 % confidence regions of length of about 0.017 for the probability of a binomial variable. The CPU-time was then between 5 and 10 minutes for each table.

## References

- IBM (1970). Application Program System/360 Scientific Package Version III Fifth Edition
- LEHMANN, E. L. (1975). Nonparametrics: Statistical methods based on ranks. Holden Day, Inc., San Francisco
- MILTON, ROY C. (1970). Rank order probabilities. John Wiley & Sons, Inc., New York, London, Sydney, Toronto
- WITTING H., NÖLLE G. (1970). Angewandte Mathematische Statistik. B. G. Teubner, Stuttgart

Anschrift des Verfassers: Dipl.-Math. Karl-Heinz Jöckel, Abteilung Statistik der Universität Dortmund, Postfach 50 05 00, 46 Dortmund 50.



## A FORTRAN program to compute exact P-values for Freeman-Halton tests

Rolf Heller

### Summary

A FORTRAN program is presented which computes exact P-values for the Freeman-Halton test. This test usually is applied if observed or expected numbers of  $n \times m$  contingency tables are too small to use the  $\chi^2$ -TEST. It coincides with the exact Fishertest for  $n=m=2$ . There are no restrictions on the program's use, with respect to either sample size or the number of rows and columns.

### Zusammenfassung

Es wird ein FORTRAN Programm vorgestellt, das exakte P-Werte für den Freeman-Halton Test berechnet. Dieser Test wird gewöhnlich angewendet, wenn die Besetzung einer  $n \times m$  Kontingenztafel so klein ist, daß sich der  $\chi^2$ -Test verbietet. Er fällt für den Fall  $n=m=2$  mit dem exakten Fisher-Test zusammen. Das Programm kann bei jeder Stichprobengröße und bei beliebiger Zahl von Reihen und Spalten angewendet werden.

Freeman and Halton (1951) presented an exact treatment of contingency tables for the cases in which expected and observed frequencies were too small for the  $\chi^2$ -test to be applied. It is not possible to give a simple answer in the form of one table because the procedure for calculating the exact P-value is depending entirely on the border totals of the contingency table.

Several authors have provided tables for special cases of  $n \times m$  contingency tables. For example Stegie and Wall (1974) tabulated P-values for  $3 \times 2$  contingency tables with sample

sizes between 6 and 15, Krüger and Wall (1975) offered P-values for the  $2 \times 3$  case where both samples have equal numbers and total sample size is between 6 and 20, whereas Bennett and Nakamura (1963) provided P-values for  $3 \times 2$  contingency tables in which the three samples are of equal size (between 3 and 20). Krüger (1976) presented tables for  $3 \times 3$  contingency tables using a test variable which determines the level of significance. Although these sources are useful and sufficient in most of the cases, a general answer to the problem is still lacking.

In this paper a FORTRAN program is presented which calculates exact P-values for arbitrary  $n \times m$  contingency tables and arbitrary sample sizes. The program first calculates the exact probability of occurrence of a given contingency table  $(x_{ij})_{i=1,\dots,n; j=1,\dots,m}$  under the null hypothesis of homogeneity assuming the border totals to be fixed. This probability is given by

$$P_L = \frac{\prod_{i=1}^n \pi(a_i!) \cdot \prod_{j=1}^m \pi(b_j!)}{N! \cdot \prod_{i=1}^n \pi_i! \cdot \prod_{j=1}^m \pi_j! (x_{ij}!)}$$

according to Freeman and Halton (1951),  $a_i$  and  $b_j$  respectively represent row and column totals and  $N$  the total number of observations. Next the probability  $P'_L$  of each combination of non-negative integers  $(x'_{ij})_{i=1,\dots,n; j=1,\dots,m}$  satisfying the border totals is calculated. Values  $P'_L$  smaller than  $P_L$  are summed up, giving the probability  $P$  of occurrence of contingency tables with equal border totals which are as likely or less likely under the null hypothesis of homogeneity. On the basis of this P-value the hypothesis can be either rejected or accepted at a given level of significance.

This program offers advantages over previous treatments of the Freeman-Halton test: a) it is applicable to larger sample sizes ( $N$ ), b) it provides P-values for uncommon combinations of  $n$  and  $m$ , and c) it gives exact P-values for any contingency table, which is advantageous in cases it is necessary to calculate a joint P-value for several independent tables.

Although the program is written for arbitrary contingency tables it should be pointed out that because the number of possibilities of combining nonnegative integers in a way satisfying the border totals increases rapidly with  $N$ ,  $n$  and  $m$ , the time needed for calculating the P-values increases correspondingly. Another

Table 1

Chimpanzee group size	Number of attacks	
	successful	unsuccessful
1	6	13
2–3	14	13
4–5	7	5
>5	4	2

Table 2

	C	B	A		C	B	A		C	B	A			
X	0	3	2	(5)	X	8	4	0	(12)	X	(8)	(7)	(2)	[17]
Y	6	5	1	(12)	Y	16	5	0	(21)	Y	(22)	(10)	(1)	[33]
	(6)	(8)	(3)	[17]		(24)	(9)	(0)	[33]		[30]	[17]	[3]	50



```

C      COMPUTATION OF EXACT P-VALUES FOR THE FREEMAN-HALTON TEST
C      N      =NUMBER OF ROWS      ;M      =NUMBER OF COLUMNS
C      X(I,J) =ACTUAL FILLING OF THE CHI-SQUARE TABLE
C      A(I)   =SUM OF ROW I      ;B(J)   =SUM OF COLUMN J
C      AA(I,J)=X(I,1)+...+X(I,J-1);BB(I,J)=X(1,J)+...+X(I-1,J)
C      INTEGER A(10),B(10),X(10,10),AA(10,10),BB(10,10),TOT
C      DOUBLE PRECISION FAKULT,Z,P,PROB,PL
      NEM=5
      NAM=6
11     READ(NEM,10,END=1000) N,M
      IS3=0
      TOT=0
      PL=0
      DO 20 I=1,N
        A(I)=0
        READ(NEM,10) (X(I,J),J=1,M)
10     FORMAT(10(I2,X))
        DO 20 K=1,M
          IF(I.EQ.1) B(K)=0
          AA(I,K)=0
          B(K)=B(K)+X(I,K)
          BB(I,K)=0
          A(I)=A(I)+X(I,K)
20     TOT=TOT+X(I,K)
        Z=1
        NN=1
        DO 24 I=1,N
          Z=Z*FAKULT(A(I))
          NK=NN+A(I)-1
          DO 25 K=NN,NK
            Z=Z/K
25     NN=NK+1
24     DO 26 J=1,M
          Z=Z*FAKULT(B(J))
26     NM1=N-1
        MM1=M-1
        LN=N-1
        LM=M-1
        GOTO 103
79     X(LN,LM)=X(LN,LM)+1
80     IF(AA(LN,LM)+X(LN,LM).GT.A(LN)) GOTO 50
        AA(LN,LM+1)=AA(LN,LM)+X(LN,LM)
        IF(BB(LN,LM)+X(LN,LM).GT.B(LM)) GOTO 50
        BB(LN+1,LM)=BB(LN,LM)+X(LN,LM)
        XS=0
        XZ=0
        DO 100 K1=1,MM1
          X(N,K1)=B(K1)-BB(N,K1)
          XZ=XZ+X(N,K1)
100     DO 101 K1=1,MM1
          X(K1,M)=A(K1)-AA(K1,M)
          XS=XS+X(K1,M)
101     IF(XZ.GT.A(N).OR.XS.GT.B(M)) GOTO 79
          X(N,M)=A(N)-XZ
103     P=Z
        DO 102 I=1,N
          DO 102 J=1,M
            P=P/FAKULT(X(I,J))
102     IF(P-PL.LE.1./10.**10) PROB=PROB+P
            IF(IS3.NE.0) GOTO 79
        DO 300 I=1,N
300    WRITE(NAM,301) (X(I,K),K=1,M),A(I)
301    FORMAT(X,10I5)
        WRITE(NAM,301) (B(K),K=1,M),TOT
        PL=P
        DO 30 I=1,N
          DO 30 J=1,M
30     X(I,J)=0
        IS3=1
        PROB=0.
        GOTO 79
50     IF(LN.EQ.1.AND.LM.EQ.1) GOTO 1
        LM=LM-1
        IF(LM.EQ.0) LN=LN-1
        IF(LM.EQ.0) LM=MM1
        X(LN,LM)=X(LN,LM)+1
        AA(LN,LM+1)=AA(LN,LM)+X(LN,LM)
        BB(LN+1,LM)=BB(LN,LM)+X(LN,LM)
        IF(AA(LN,LM+1).GT.A(LN)) GOTO 50
        IF(BB(LN+1,LM).GT.B(LM)) GOTO 50
        DO 60 K1=LN,MM1
          DO 60 K2=1,MM1
            IF(K1.EQ.LN.AND.K2.LE.LM) GOTO 60
            X(K1,K2)=0
            AA(K1,K2+1)=AA(K1,K2)
            BB(K1+1,K2)=BB(K1,K2)
60     CONTINUE
        LN=N-1
        LM=M-1
        GOTO 80
1     WRITE(NAM,400) PROB
400    FORMAT(' P-VALUE = ',D20,10)
1000   GOTO 11
      STOP
      END
      DOUBLE PRECISION FUNCTION FAKULT(N)
      FAKULT=1
      IF(N.LE.1) RETURN
      DO 10 I=1,N
10     FAKULT=FAKULT*I
      RETURN
      END

```

problem could be the occurrence of large numbers (e. g.  $N!$ ) inducing an overflow. However, the program presented here minimizes this risk by using double precision variables.

### Examples

a) Busse (1978) compared the hunting success of chimpanzees in groups of different sizes on red colobus monkeys (Table 1). Although there is a slight increase of hunting success with increasing group size, no significant differences are found by the Freeman-Halton test ( $P=0.321 > 0.05$ ).

b) Freeman and Halton (1951) provide three further examples which are presented in Table 2 »The sample consists of fifty candidates who passed successfully through a training establishment having previously passed successfully through a selection centre. The candidates are classified by selection centre grade A, B, or C, by training establishment grade X and Y and also by an educational factor M and N (Totals are given in parentheses, totals of totals in brackets and the overall figure is in black.) < P-Values are 0.088, 0.690, and 0.225. Thus, at the 5% level, there is no evidence of heterogeneity in the distribution of selection centre and training establishment grades in the cases of candidates with educational factor M, N and for pooled data.

The output of the program for these examples is added to test copied programs.

a)	6	13	19	
	14	13	27	
	7	5	12	
	4	2	6	
	31	33	64	
P-VALUE =	0.3207216023D+00			
b)	0	3	2	5
	6	5	1	12
	6	8	3	17
P-VALUE =	0.8823529412D-01			
	8	4	0	12
	16	5	0	21
	24	9	0	33
P-VALUE =	0.6904594849D+00			
	8	7	2	17
	22	10	1	33
	30	17	3	50
P-VALUE =	0.2250563046D+00			

### References

- BENNETT, B. M. and NAKAMURA, E. (1963): Tables for testing significance in a 2 x 3 contingency table. *Technometrics* 5, 501-511
- BUSSE, C. D. (1978): Do chimpanzees hunt cooperatively? *American Naturalist* 112, 767-770
- FREEMAN, G. H. and HALTON, J. H. (1951): Note on an exact treatment of contingency, goodness of fit and other problems of significance. *Biometrika* 38, 141-149
- KRÜGER, H. P. (1976): Tafeln für einen exakten 3 x 3-Felder Kontingenztest. EDV in Biologie und Medizin (im Druck).
- KRÜGER, H. P. and WALL, K. D. (1975): Exakte Homogenitätsprüfung in 2 x 3-Feldertafeln mit zwei Stichproben gleichen Umfangs. EDV in Biologie und Medizin (im Druck)
- STEGIE, R. and WALL, K. D. (1974): Tabellen für den exakten Test in 3 x 2-Felder-Tafeln. EDV in Biologie und Medizin 5, 73-82

Anschrift des Verfassers: Rolf Heller, Arbeitsgruppe für Verhaltensforschung, Abteilung für Biologie, Ruhr-Universität Bochum, Postfach 10 21 48, 4630 Bochum

Listing of the FORTRAN program



## BUCHBESPRECHUNGEN

MÜLLER, C. D.

### Besondere Vertragsbedingungen für die Überlassung von DV-Programmen (BVB DV-Programm-Überlassung)

1978, 146 S., DM 42,-  
E. Schmidt Verlag, Berlin – Bielefeld – München

Da die Softwarekosten in die Größenordnung der Hardwarekosten gekommen sind, ist es verständlich, daß auch die Überlassung von DV-Programmen geregelt wird.

Dem Text der »Besonderen Vertragsbedingungen« folgen in dem vorliegenden Leitfaden die paragrafenweisen Erläuterungen zu den einzelnen Bestimmungen. Sie geben gerade demjenigen Anwender, der sich nicht ständig mit der Vertragsgestaltung im EDV-Wesen befaßt, eine Hilfe beim Verständnis der BVB und bieten ihm Anregungen für die Beurteilung und Gestaltung anderer Vertragsmuster.

Mit abgedruckt sind die Formularmuster und die VOL/B, da ohne sie die »Besonderen Vertragsbedingungen« nicht verstanden werden können.

Ge

KRES, H.

### Statistische Tafeln zur multivariaten Analysis

1975, 431 S., DM 48,-  
Springer-Verlag, Berlin – Heidelberg – New York

Die multivariate Analysis wird immer mehr angewandt. Daß dabei die Probleme nicht einfacher werden, zeigt sich insbesondere auch bei der Erarbeitung von Teststatistiken.

Es ist das Verdienst des Verfassers, daß er eine Vielzahl sehr verstreut veröffentlichter Tabellen einmal zusammengestellt hat. Wenn auch die Behandlung der Methoden und die Anwendungen weiteren Darstellungen vorbehalten bleiben, so muß doch an dieser Stelle die klare, überschaubare und gut strukturierte Darstellung lobend erwähnt werden.

Die Anwender haben mit diesem Buch ein Hilfsmittel, das ihnen bei der Bearbeitung multivariater Fragestellungen sehr nützlich sein wird.

Ge.

EIMEREN, W. VAN (Hrsg.)

### Perspektiven der Gesundheitssystemforschung

Medizinische Informatik und Statistik Bd. 10 1978, 171 S., DM 27,-  
Springer-Verlag, Berlin – Heidelberg – New York

Eigentlich muß man fragen, warum erst so spät die Methoden der Systemforschung auf Probleme des Gesundheitswesens angewandt werden. Auf jeden Fall ist es zu begrüßen, daß nun ein interdisziplinärer Gedankenaustausch beginnt. Der vorliegende Band enthält die Referate einer ersten Tagung (Frühjahr 1978) des Fachbereichs »Planung und Auswertung« der GMDS zu diesem Thema. Interessant ist dabei die Vielfalt der Aspekte, die in z. T. ausgezeichneten Referaten behandelt wurden.

Ge.

SCHINDLER, S. und GILOI, W. K. (Hrsg.)

### GI – 8. Jahrestagung

Informatik-Fachberichte Bd. 16  
1978, 394 S., DM 35,50

Springer-Verlag, Berlin – Heidelberg – New York

Die GI-Jahrestagung 1978 war nicht wie bisher speziellen Fachthemen gewidmet. Im Mittelpunkt standen Fachgespräche und Diskussionen um die gesellschaftlichen und wirtschaftlichen Implikationen der neuen Informatik-Technologien. Die wesentlichsten Beiträge sind in dem vorliegenden Tagungsband zusammengefaßt und geben damit einen guten Überblick über den augenblicklichen Stand der Diskussionen.

Ge.

## BUNDESGESUNDHEITSAMT



Das Bundesgesundheitsamt ist ein Bundesoberbehörde. Zu seinen Aufgaben gehören anwendungsorientierte Forschung, Beratungs- und Exekutivaufgaben, die Erstellung von Gutachten sowie medizinisch-statistische Arbeiten auf dem Gebiet des öffentlichen Gesundheitswesens. Ziele sind die Verbesserung des gesundheitlichen Verbraucherschutzes, die Verminderung gesundheitlicher Risiken aus der Umwelt, die Erforschung von Krankheitsursachen und die Früherkennung von Krankheiten, die Krankheitsbekämpfung und die Arzneimittelsicherheit. Es beschäftigt ca. 1500 Mitarbeiter, davon 400 Wissenschaftler.

Im Institut für Strahlenhygiene in NEUHERBERG sind folgende Stellen zu besetzen:

**Kenn-Nr. 20/79-V 1-**

Ab sofort 1 Stelle (zunächst befristet bis 1982) als

### Wissenschaftliche(n) Angestellte(n)

(Verg. Gr. II a BAT)

Die Übernahme auf unbestimmte Zeit ist bei Bewährung möglich.

**Aufgabengebiet:** Entwicklung und Anwendung von Methoden zur Abschätzung von Risiken der Bevölkerung durch Belastungen aus der Anwendung radioaktiver Stoffe und ionisierender Strahlung. Erfassung, Dokumentation und Verarbeitung von strahlenexpositionsrelevanten Daten mit Methoden der Datenverarbeitung und Statistik.

**Anforderungen:** Statistiker oder Naturwissenschaftler mit Kenntnissen in der angewandten Statistik. Praktische Erfahrung im Umgang mit elektronischen Datenverarbeitungssystemen ist erforderlich. Der Bewerber muß befähigt sein, in dem interdisziplinären Arbeitsgebiet selbständig wissenschaftlich zu arbeiten.

Bewerbungen mit Lebenslauf, Lichtbild und kurzer Darstellung des beruflichen Werdeganges sowie Zeugnisabschriften und gegebenenfalls einem Verzeichnis der Veröffentlichungen sind unter Angabe des für Sie in Frage kommenden Arbeitsgebietes und der entsprechenden Kenn-Nummer bis zum 15. August 1979 an das

**Bundesgesundheitsamt**

– Personalreferat –

**Postfach, 1000 Berlin 33,**  
zu richten.

Das europäische  
EDV-Forum

# SYSTEMS'79

**SYS**

Computersysteme und Ihre  
Anwendung

- Internationale Seminare für EDV-Anwendung und -Ausbildung, Systementwicklung, Normung, OEM, OCR
- Internationale Fachausstellung der EDV- und Softwareindustrie

Zusammen mit

**5. Europäischem**

**Mikrofilm-Kongreß**

und Fachausstellung

München, 17. bis 21. September 1979

Fordern Sie Informationen an

Münchener Messe- und Ausstellungsges. mbH

Postfach 121009, 8000 München 12, Tel. 089/5107-1